

НАРОДНА УКРАЇНЬСЬКА АКАДЕМІЯ

О. В. Лазаренко, Д. І. Панченко

**ДОСЛІДЖЕННЯ
СЕМАНТИЧНИХ ЗВ'ЯЗКІВ
ТЕКСТ-РЕФЕРАТ
ДЛЯ СТВОРЕННЯ СИСТЕМИ
АВТОМАТИЧНОГО
РЕФЕРУВАННЯ
З ОПЕРТЯМ НА ЗНАННЯ**



Харків
Видавництво НУА
2013

УДК 81'322.2'33
ББК 81.11
Л 17

*Затверджено на засіданні кафедри інформаційних технологій
та математики
Народної української академії
Протокол № 6 від 21.01.2013*

Рецензенти: *д-р техн. наук К. О. Соловйова (ХНУРЕ)
д-р техн. наук Н. В. Шаронова (НТУ «ХПІ»)*

Лазаренко, Ольга Володимирівна.

Л 17 Дослідження семантичних зв'язків ТЕКСТ-РЕФЕРАТ для створення системи автоматичного реферування з опертям на знання : монографія / О. В. Лазаренко, Д. І. Панченко ; Нар. укр. акад., [каф. інформ. технологій та математики]. – Х. : Вид-во НУА, 2013. – 180 с.

ISBN

У виданні досліджуються питання побудови семантико-контекстної моделі реферування та моделі подання знань у системі автоматичного реферування з використанням онтологій. Розглянуто семантичні зв'язки тексту, заголовка та реферату, їх специфіка й особливості. Розроблено процедуру автоматичного здобування знань із текстів природною мовою з використанням моделі реферату, моделі заголовка та семантико-контекстної моделі реферування, що дозволило наблизитися до формального опису певних аспектів процесу розуміння тексту людиною.

УДК 81'322.2'33
ББК 81.11

ISBN

© Народна українська академія, 2013

ВСТУП

Підвищення якості автоматичної обробки текстової інформації постійно перебуває в центрі уваги розробників систем перекладу, реферування й пошуку інформації. На сьогодні роботи в цих галузях упритул підійшли до необхідності розробки семантичних технологій, спрямованих на автоматичне розпізнавання змісту текстів. Розробка таких технологій є одним із найскладніших завдань при проектуванні інтелектуальних систем. Окреме місце в цьому ряду займають системи, орієнтовані на автоматичне реферування текстів – побудову стислих, змістовних, виразних фрагментів, які в основному відображають зміст реферованих текстів. Роботи у сфері створення систем автоматичного реферування мають багате минуле. Понад півстоліття відбувались активні дослідження, присвячені пошукам ефективних методів автоматичного реферування (АР). Найбільш суттєвими вважаються класичний статистичний (Г. П. Лун, Р. Г. Піотровський) і позиційний (М. Л. Лосєва, М. С. Максудова) методи. Започатковувалися вдалі спроби дослідження лінгвістичних і структурних характеристик реферату (В. І. Горькова, Е. А. Борохов, Х. Борко, С. Берньє, Р. Мейзел, Дж. Сміт), проводилися дослідження з морфологічного й синтаксичного аналізу текстів (В. С. Перебийніс, Н. Ф. Клименко, Г. Е. Мірам, Т. О. Грязнухіна, Н. П. Дарчук та ін.), формалізації понадфразових зв'язків у оригінальному тексті (І. П. Севбо, Л. В. Орлова) і використання інформаційно-логічних мов для побудови реферату автоматичним шляхом (Е. Ф. Скороходько).

Усі спроби автоматизації реферування, до яких удавалися протягом останніх років, одержали однозначний підсумок – залучення виключно статистичних і позиційних методів не дозволяє створити повноцінну систему реферування, яка б замість вилучення речень із тексту стисло викладала зміст. Тому протягом останніх років увага дослідників зосереджувалася переважно на аналізі змісту текстової інформації (Г. С. Жданова, Г. Г. Белоногов, Р. Г. Котов, І. Мейні, М. Мейбюрі та ін.). Для аналізу великих масивів неструктурованої

інформації використовують технології Text Mining, цікаві спроможністю здобуття високоякісної інформації з тексту, аналізу тексту, який об'єднує в собі технологічні й методологічні підходи до опрацювання змісту. Програми, які розв'язують ці завдання, повинні певним чином розуміти семантику аналізованого тексту. Тут дослідження спираються на досягнення структурної і прикладної лінгвістики, когнітивної семантики, логічної семантики, математичної логіки та ряду інших фундаментальних і прикладних дисциплін. Результати цих теоретичних розробок знаходять застосування в системі автоматизації процесів опрацювання інформації та побудові інтелектуальних інформаційних систем у різних галузях науки (Д. В. Ланде, В. А. Широков).

У сфері інформаційних технологій позначився позитивний перехід до розробок, зосереджених на знаннях як якісних складових таких процесів. Розв'язувана в даному дослідженні проблема моделювання семантичних зв'язків «Текст-Реферат» у системах автоматичного реферування (АР) зосереджена на створенні системи АР з опертям на знання й передбачає змістове опрацювання тексту в автоматичному режимі.

При цьому базовою моделлю опису знань при побудові систем АР обрано *онтології* як засіб класифікації терміносфери предметної галузі. Останні дозволяють автоматизовано опрацювати семантичну інформацію. Вказані можливості онтологій зумовлюють актуальність розробки останніх як засобу впровадження технологій, що уможливають розвинене машинне опрацювання інформації шляхом здійснення змістового аналізу текстів.

Однак змістовий аналіз тексту є лише одним з етапів процесу реферування, після якого йдуть не менш складні завдання – стиснення змісту й синтез реферату. Отже, пошук шляхів і методів автоматичної компресії тексту є, на наш погляд, дуже важливою складовою досліджень при розробці систем автоматичного реферування. Під *компресією* розуміємо сукупність операцій аналітико-синтетичного опрацювання інформації, які мають на меті створення

вторинних документів чи демонстрацію змісту вихідного тексту в більш економній формі при максимальному збереженні його інформативності в похідному тексті. Реалізація механізмів компресії у системі АР дозволяє істотно прискорити й полегшити процес укладання рефератів і сприяє інтелектуалізації процесу реферування.

Сказане вмотивовує актуальність дослідження моделювання семантичних зв'язків у природно мовних текстах з метою побудови ефективних систем автоматичного реферування.

У цьому контексті метою нашого дослідження стала побудова семантико-контекстної моделі реферування та моделі подання знань у системі АР із використанням онтологій з метою оптимізації процедури автоматичного аналізу та компресії текстів у процесі реферування.

Для досягнення цієї мети необхідно було розв'язати, перш за все, такі питання:

- розробити методику формування ознак, що характеризують реферат і заголовки як результат компресії інформації на різних рівнях процесу реферування;

- провести дослідження синтаксичної структури реферативних конструкцій та первинних текстів і побудувати модель процесу згортання на синтаксичному рівні;

- провести аналіз процесу згортання на семантико-синтаксичному, власне семантичному й лексико-семантичному рівнях і побудувати модель процесу згортання на семантичному рівні;

- побудувати модель реферату у вигляді типових для індикативних рефератів семантико-синтаксичних структур;

- побудувати модель заголовка, що відображає подібність його з рефератом, з метою більш ефективного використання цієї моделі в процесі автоматичного реферування;

- побудувати текстову базу, що відображає змістове ядро тексту;
- провести класифікацію лексики, яка бере участь у формуванні індикативних рефератів і заголовка;
- створити онтології на базі укладених словників: онтологію верхнього рівня та виродженої онтологію загальнонаукової лексики.

Для вирішення цих завдань об'єктом дослідження стали семантичні зв'язки тексту й реферату, їх специфіка й особливості. В рамках об'єкта було розглянуто процедури автоматичного здобування знань із текстів природною мовою з використанням моделі реферату, моделі заголовка, текстової бази та онтологій, що стало безпосередньо предметом проведеного дослідження.

Вихідним пунктом нашого дослідження стало положення про те, що моделювання процесу реферування як сукупності найскладніших процесів розуміння й компресії змісту слід починати з вивчення не самих процесів, а з їх результату – реферату. Причому не розгорнутого, інформативного, а стиснутого, індикативного, перш за все, тому що розглядаємо його як відправну точку в дослідженні цього питання, як об'єкт найбільш простий за формою, але такий, що відображає всі особливості реферативного тексту. Передбачалось, що дослідження значеннєвої і синтаксичної структур реферату дозволить з'ясувати природу компресії в реферуванні та її можливі наслідки щодо структури реферативних речень, і на підставі виявлених особливостей семантико-синтаксичної структури цих речень можливо буде побудувати модель індикативного реферату. Наступним кроком передбачався перехід до розробки процедури здобування знань із тексту й заповнення моделі реферату відібраними з тексту іменниковими групами. При цьому відправною точкою для змістового аналізу тексту розглянуто заголовки (назва статті), значеннєва й синтаксична структури якого дозволяють знайти найважливіші змістові елементи для пошуку іменникових груп у тексті й побудови на їх основі текстової бази.

Матеріалом дослідження було обрано реферативні конструкції, які будують тексти індикативних рефератів. Кількість рефератів, що зібрано методом суцільної вибірки з текстів чотирьох предметних галузей, становитиме: економіка (1000 рефератів), медицина (1000 рефератів), прикладна лінгвістика (2000 рефератів) та соціоніка (1000 рефератів). Загалом проаналізовано лінгвістичні характеристики приблизно 5000 реферативних конструкцій та первинні тексти, що їм відповідають.

Методологічною основою дослідження є когнітивний підхід до аналізу текстів рефератів та їх першоджерел. На основі когнітивного підходу побудована модель процесу реферування з опертям на знання. Теоретична база досліджень – когнітивна лінгвістика, зокрема методи когнітивної семантики, й теорія вербоцентричного синтаксису залежностей. Завдання дослідження визначили конкретні методи аналізу й класифікації, зокрема когнітивний (концептуалізація та категоризація), позиційний, статистичний, експериментальний.

Наукова новизна отриманих результатів полягає в:

- розробці методики формування ознак, що характеризують реферат як результат компресії інформації на різних рівнях процесу реферування, та побудові моделі реферату науково-технічного тексту;
- побудові моделі заголовка й дослідження зв'язку заголовка з моделлю реферату;
- розробленні процедури добування з тексту знань, необхідних для заповнення моделі реферату;
- побудові моделі семантичних зв'язків «текст-реферат» (семантико-контекстної моделі реферування) з використанням текстової бази, онтологій двох рівнів і змістових елементів заголовка;
- створенні та опису онтологій верхнього рівня та загальнонаукової лексики;

- удосконаленні алгоритму заповнення реферативних конструкцій у моделі реферату іменниковими групами й предикативними елементами з тексту за допомогою заголовка, текстової бази та онтологій.

Теоретичне значення виконаного дослідження полягає в розробці когнітивної (або семантико-контекстної) моделі реферування, що дозволило зробити висновки про те, як відбувається змістове згортання в процесі реферування, й у припущенні, що побудова текстової бази, яка являє собою основу макроструктури тексту, дозволить наблизитися до формального опису певних аспектів процесу розуміння тексту людиною.

Отримані результати дозволили з'ясувати, як відбувається змістове згортання в процесі реферування і які специфічні ознаки у структурі реферативних речень і заголовку воно унаочнює. Розроблені моделі орієнтовані на використання в системах автоматичного опрацювання інформації для покращення якості процесів реферування.

Теоретичні результати досліджень використані в навчальному процесі на кафедрі інформаційних технологій та математики при викладанні спеціальних дисциплін «Інформаційні технології референта-перекладача» та «Сучасні методи автоматичного опрацювання інформації» у ХГУ «НУА».

І. ОГЛЯД СУЧАСНИХ ДОСЛІДЖЕНЬ У ГАЛУЗІ АВТОМАТИЧНОГО РЕФЕРУВАННЯ

1. 1. Розвиток ідеї автоматичного реферування

Проблема автоматичного опрацювання інформації (АОІ) набуває з кожним роком усе більшого значення й ваги. Це зумовлено стрімким зростанням обсягів документів на електронних носіях у всіх сферах життя сучасного суспільства й невідпинними темпами розвитку всесвітньої мережі Інтернет. У інформаційних сховищах, розповсюджених по всьому світу, нагромаджені і продовжують накопичуватися терабайти текстових даних. Знайти в таких масивах необхідні й корисні відомості можливо лише за сприянням спеціалізованих технологій. За наявними оцінками лише 10% усього масиву інформації, з яким мають справу користувачі, припадає на структуровані дані, решта – 90% – неструктуровані [105, ст. 159].

Тому одне з найважливіших і першочергових завдань у опрацюванні текстової інформації полягає в пошуку шляхів і методів автоматичного стиснення тексту, скорочення його обсягу при збереженні основного змісту, тобто побудова рефератів і анотацій.

Серед технологій та систем, спрямованих на врегулювання інформаційних потоків, важливе місце посідають інтелектуальні системи (ІС) подання знань. На початку свого становлення й розвитку інформаційні системи функціонували лише на рівні опрацювання даних і тексту документів, згодом, завдяки прогресу в галузі інформаційних технологій, з'явилися системи, які намагаються працювати вже на змістовому (понятійному) рівні.

ІС застосовуються там, де основну складність представляють неформалізовані знання спеціалістів-практиків і де змістове опрацювання інформації превалює над обчислювальним. Під час проектування такого типу систем, на думку Т. А. Гаврилової і В. Ф. Хорошевського, слід враховувати

складність предметних галузей та управління процесом розробки, гнучкість кінцевого програмного продукту та описання поведінки окремих підсистем [41, ст. 91].

Більш поширеним класом інтелектуальних систем відгалужуються експертні системи, які акумулюють знання фахівців у конкретних предметних галузях, де якість прийняття рішень залежить від рівня експертизи (медицина, юриспруденція, геологія, економіка, військова справа тощо). Однак вони ефективні тільки в «експертних» галузях, де важливим є емпіричний досвід спеціалістів.

На сьогодні, на наш погляд, не існує універсальних інтелектуальних систем, зручних для широкого кола користувачів при роботі в мережі Інтернет. Щодо наявних систем, то всім їм тією чи іншою мірою бракує потрібних можливостей, а саме:

- 1) наявності внутрішньої моделі зовнішнього світу;
- 2) здатності поповнення знань;
- 3) спроможності до дедуктивного висновку, тобто генерації інформації, яка в явному вигляді не міститься у системі;
- 4) уміння оперувати в ситуаціях із різними аспектами нечіткості, включно з «розумінням» природної мови;
- 5) здатності до діалогової взаємодії з людиною;
- 6) адаптації до різних умов роботи.

З огляду на це актуальною залишається розробка теорії, методів і технологій подання й використання знань для створення й запровадження інтелектуальних систем.

До таких систем можна віднести системи автоматичного реферування (АР), покликані зменшити напругу при роботі з численними інформаційними масивами, які надали б можливість ознайомлення зі стислим змістом документів, суттєво полегшили роботу з інформацією. Запропоновані програмні продукти, на

жаль, не здатні забезпечити оперативне отримання рефератів і анотацій необхідної повноти й релевантності. Тому розробка якісних автоматичних рефератів залишається вкрай необхідною і водночас є найскладнішим завданням із машинного опрацювання інформації.

Проблема автоматизації процесу реферування постала не сьогодні і має свою історію [68]. Небувалого імпульсу вона набула з появою Інтернету. Сьогодні налічується чимало методів і систем AP, основні з яких описані в підрозділах 1.2 і 1.3.

Еволюціонували й думки фахівців із цього приводу. Вважалося, що реферат здатний повністю замінити читачеві першоджерело. Протилежний погляд – реферат лише дає уявлення про необхідність звернення до першоджерела [21]. У першому випадку акцентувалися інформативні функції реферату, а в другому – індикативні. З розмежуванням обох функцій це протиріччя зникає. Відповідно ці функції виконуються інформативними та індикативними рефератами.

Не менш палкі суперечки точилися й тривають навколо розмежування понять «реферат» і «анотація». Підґрунтям полеміки є відносність розподілу в тексті першоджерела власне інформації, носієм якої є реферат, і метainформації, носієм якої є анотація, а також нерозуміння принципової різниці між даними видами вторинних документів [209; 124, ст. 222–223].

А. А. Вейзе під анотацією розуміє дуже стислий виклад змісту первинного документа, що дає загальне уявлення про тему [33, ст. 66]. В цьому, на його думку, полягає один із суттєвих моментів відмінності анотації від реферату, який у стислій формі знайомить читача з сутністю змісту оригіналу.

На думку А. І. Михайлова, «в анотуванні головним є вміння лаконічно *стиснути* зміст документа, реферування передбачає володіння мистецтвом *скорочення* тексту первинного документа» [124, ст. 224].

Ми дотримуємося думки, згідно з якою *реферат* передає фактографічну інформацію і відповідає на питання, яка інформація закладена в первинному документі (формулювання та узагальнення запозичуються з тексту оригіналу); *анотація* ж є стислою описовою характеристикою першоджерела й відповідає на запитання, про що говориться в первинному документі (головний зміст передається своїми словами, абстраговано та узагальнено). До цього слід додати, що анотація орієнтує користувача на інформацію в документальних потоках (виконує сигнальну функцію), а реферат – в інформаційних (пізнавальна функція).

Порівняно з анотуванням реферування є більш розвиненим методом опрацювання і більш зручним для автоматизованого подання інформації.

Авторами використовуються різні визначення поняття «реферат».

– Реферат (від лат. *referre* – повідомляти, доповідати) – це короткий виклад змісту наукового документа [174].

– Реферат – модель оригіналу, що пропорційно відображає його складники [70, ст. 24–25].

– Реферат – стислий, точний виклад змісту документа, що містить головні фактичні відомості та висновки без додаткової інтерпретації або критичних зауважень автора реферату, в якому акцентується увага на нових відомостях і визначається доцільність звернення до першоджерела, а формулювання та узагальнення запозичуються з тексту оригіналу [62, ст. 132].

– Реферат – це семантично адекватний, обмежений за обсягом і водночас за можливістю повний виклад основного змісту первинного документа, що характеризується сталою структурою і призначений для виконання різноманітних інформаційних функцій при використанні його читачами різних категорій [55].

Усі дефініції мають, на наш погляд, певний сенс, утім, аби уникнути неоднозначності трактування ключового для цієї роботи поняття, ми будемо дотримуватися останнього з наведених тлумачень.

Структура реферату визначається ступенем важливості відібраних відомостей: спочатку в концентрованій формі викладається зміст питання, далі, якщо є потреба, коротко наводяться необхідні фактичні дані.

Призначення реферату – швидке, систематичне подання актуальної науково-технічної інформації у згорнутому вигляді на основі її змістового опрацювання [78, ст. 62]. Тому головною жанровою рисою мови реферату є інформативність, яку слід розглядати як на лексичному, так і на синтаксичному рівні [123].

Чільною ланкою застосування систем автоматичного реферування стали інформаційно-пошукові системи (ІПС). Розвиток обчислювальної техніки та комп'ютерних мереж сприяв появі систем, які здійснюють пошук у масивах повнотекстових документів. До певного часу ІПС використовувалися здебільшого фахівцями, коло яких було обмежене – архівні працівники, співробітники бібліотек, науковці, аналітики. З появою і поширенням мережі Інтернет ситуація змінюється докорінно. Сьогодні інформаційні ресурси налічують значну кількість Web-сторінок, доступних будь якому користувачеві. Знайти необхідну інформацію в повнотекстовій базі даних – справа копітка й марудна, навіть за умови залучення потужних пошукових систем, що опанували новітній ринок інформаційних технологій (Google, Yahoo, AllTheWeb, AltaVista, Yandex).

Уже сьогодні багато сучасних мережевих пошукових систем мають такі компоненти:

- автоматичне групування документів за визначеним класифікатором;
- автоматичне визначення нових класів на основі неструктурованих документів;
- ранжування документів за змістовою релевантністю;
- виявлення семантично подібних документів за еталоном;
- автоматичний аналіз і змістове перетворення запитів користувача.

Саме середовище надходження інформації визначає можливі підходи до її опрацювання. Для вирішення проблеми інформаційного хаосу найбільш прийнятні технології Data Mining і Text Mining.

Провідний світовий експерт із GTE Labs Григорій П'ятецький-Шапіро формулює Data Mining як «процес виявлення в сирих даних раніше невідомих, нетривіальних, практично необхідних і доступних інтерпретацій знань, які необхідні для прийняття рішення в різних галузях людської діяльності». Технології Text Mining цікаві спроможністю здобуття високоякісної інформації з тексту, аналізу тексту, який поєднує в собі технологічні й методологічні підходи аналізу змісту, комп'ютерної лінгвістики та штучного інтелекту або ж алгоритмічним виявленням раніше невідомих зв'язків і кореляцій у вже наявних текстових даних.

Слід зауважити, що технології «глибинного аналізу тексту» історично передували технології «глибинного аналізу даних», методологія і підходи якої широко використовуються в методах Text Mining.

На відміну від Data Mining, яка оперує однорідними даними декількох типів (покупці – товари, історія торгів – ціни на акції) і передбачає аналіз упорядкованої в певні структури інформації, технологія Text Mining аналізує великі масиви неструктурованої інформації. Програми, які розв'язують це завдання, повинні певним чином оперувати людською мовою і при цьому розуміти семантику тексту, що аналізується.

Напрямок Text Mining включив у себе всі чинні практичні дослідження контент-аналізу й комп'ютерної лінгвістики, яка подібно до теорії баз знань інтенсивно розвивалася в 70–80-ті роки минулого століття. Практично всі відомі виробники програмного забезпечення пропонують на ринку системи глибинного аналізу даних і текстів (наприклад, компанія Oracle – Oracle Text, IBM – Intelligent Miner for Text, SAS – Text Miner).

У середині 90-х років ХХ століття дана технологія як напрямок аналізу неструктурованих текстів взяла на озброєння методи класичного здобуття даних (класифікація або кластеризація), однак з'явилися і додаткові можливості: автоматичне реферування, виділення понять, виявлення змістових взаємозв'язків окремих фрагментів у тексті, відповіді на запити, пошук за ключовими словами, створення таксономії і тезаурусів, прогнозування й аналіз відхилень.

На сьогодні, на думку Д. В. Ланде, Text Mining представляє новий взірць свідомого пошуку, який, на відміну від традиційних підходів більшості когнітивних технологій, не тільки знаходить списки документів, формально релевантних запитам, а й допомагає розібратися з віднайденою інформацією [105, ст. 160].

У загальному плані автоматичне реферування тісно пов'язане з проблемою згортання інформації. Не занурюючись у детальний розгляд цього питання, обмежимося лише його прикладною складовою, розуміючи під цим можливість автоматичного перетворення текстового документа в зручну форму для подальшого машинного опрацювання.

Завдання автоматичного укладання рефератів пов'язане з завданням автоматичного індексування повідомлень (у простому випадку – індексування заголовків документів), а також з проблемою автоматичного кодування текстових повідомлень. На практиці пошукові реферати часто є тотожними наборами індексів документів [9, ст. 114].

Залежно від повноти й форми викладу змісту первинного документа, від змістових і формальних характеристик, від цільової та адресної спрямованості реферати поділяються на *інформативні* (реферат-конспект), *індикативні* (реферат-резюме), *критичні, монографічні, зведені, аспектні, фрагментні, цільові, загальні, анкетні (табличні)* [9, ст. 112–113; 174; 108; 201; 189; 214].

Концепція типології рефератів започаткована ще в 1949 р. (Париж, Міжнародна конференція з наукового реферування). В подальшому вона

удосконалювалася, розвивалася вітчизняними та зарубіжними фахівцями. Із вітчизняних науковців заслуговують на увагу вичерпні, багаті на емпіричний матеріал дослідження В. І. Соловйова [173]. Із зарубіжних авторів певний інтерес становлять праці Х. Борко, С. Бернье [214], Р. Мейзел [262], Дж. Сміт [279, ст. 32–40] та ін.

Одержання рефератів будь-якого типу пов'язане з процедурою автоматичного вилучення з тексту певних фрагментів, які можуть розглядатися як «носії змісту», документ або слугують первинним матеріалом для подальшого опрацювання. Тому автоматичне реферування власне є різновидом пошуку інформації в масивах неформалізованих документів. Звідси і складність від пошуку у масивах формалізованих документів.

У нашому дослідженні ми маємо справу зі спеціалізованими індикативними рефератами, які містять положення, тісно пов'язані з темою реферованого документа, все другорядне, для даної теми, не розглядається. На нашу думку, саме даний вид реферату викликає інтерес при документальному пошуку й використовується під час прийняття рішення про необхідність перегляду першоджерела. Крім того, в рефераті застосовуються формулювання узагальнення, запозичені з тексту оригіналу, що робить більш реальним створення автоматичних рефератів.

Основні вимоги до якісного автоматичного реферату мало чим відрізняються від вимог до інтелектуального, тобто створеного людиною, реферату. На думку Ю. І. Шемакіна, застосування рефератів як пошукових описів першоджерела і як вторинного документа, спроможного замінити оригінал, ставить цілу низку специфічних вимог до змісту, структури та мови [197, ст. 101–114; 199, ст. 100]. Серед них:

1) стислий, точний та об'єктивний виклад матеріалу, релевантність першоджерелу;

2) застосування стандартної термінології, що виключає можливість неоднозначного сприйняття термінів, які використовуються в різних аспектах однієї проблеми;

3) складні речення перетворюються в прості, широко використовуються неозначено-особові речення, тенденція до субстантивації;

4) формули, ілюстрації і таблиці в рефераті подаються у випадках, коли без них не можлива побудова реферату (відображають підсумки роботи або суттєво полегшують сприйняття).

Призначення реферату зумовлюється конкретними інформаційними потребами в діапазоні «індикативність – інформативність» залежно від інтересів користувача до предмету джерела, що реферується. Також В. П. Леонов та В. І. Соловйов додають адресну (бібліографічний опис), довідкову, комунікативну, оцінювальну, сигнальну, прогностичну функції [108; 174].

Головне покликання реферату як «інформаційної моделі», на думку А. А. Вейзе, – швидке, систематичне подання актуальної науково-технічної інформації у згорнутому вигляді на базі її змістового опрацювання [33, ст. 62–63].

У системі наукової комунікації реферат виступає як основна інформаційно-комунікативна одиниця, що зумовлюється такими його властивостями [55; 97, ст. 28–32; 186]:

1) з усіх видів вторинних інформаційних документів (анотації, резюме, бібліографічного опису) реферат вирізняється найбільшою інформативністю в розкритті змісту першоджерела;

2) використання рефератів для пошуку поточної або ретроспективної інформації може економити час, необхідний на випадок звернення до первинних документів;

3) форма надання інформації у вигляді реферату зручна для зберігання у фондах довідково-інформаційних служб, у тому числі в автоматизованих інформаційних бібліотечних системах, а також у ПС.

Завдання автоматичного реферування текстових повідомлень актуальне, оскільки укладання рефератів уручну в автоматизованих ППС є надзвичайно трудомістким процесом. Складність формалізації тут обумовлена не тільки необхідністю розроблення методів вилучення з тексту одиниць змісту, але й необхідністю їх порівняльної оцінки, аби якоюсь мірою компенсувати відсутність фахової людини-референта з її знаннями, досвідом і вмінням виокремити головний зміст документу [9, ст. 112]. Тому розв'язання завдання автоматичного укладання рефератів безпосередньо залежить від рівня формалізації цього процесу в цілому.

У роботі І. Мейні, М. Мейбюрі алгоритм автоматичного реферування передбачає три послідовних етапи: аналіз вихідного тексту, визначення вагомих фрагментів (речень або цілих абзаців) і формулювання реферату (рис. 1.1) [263].

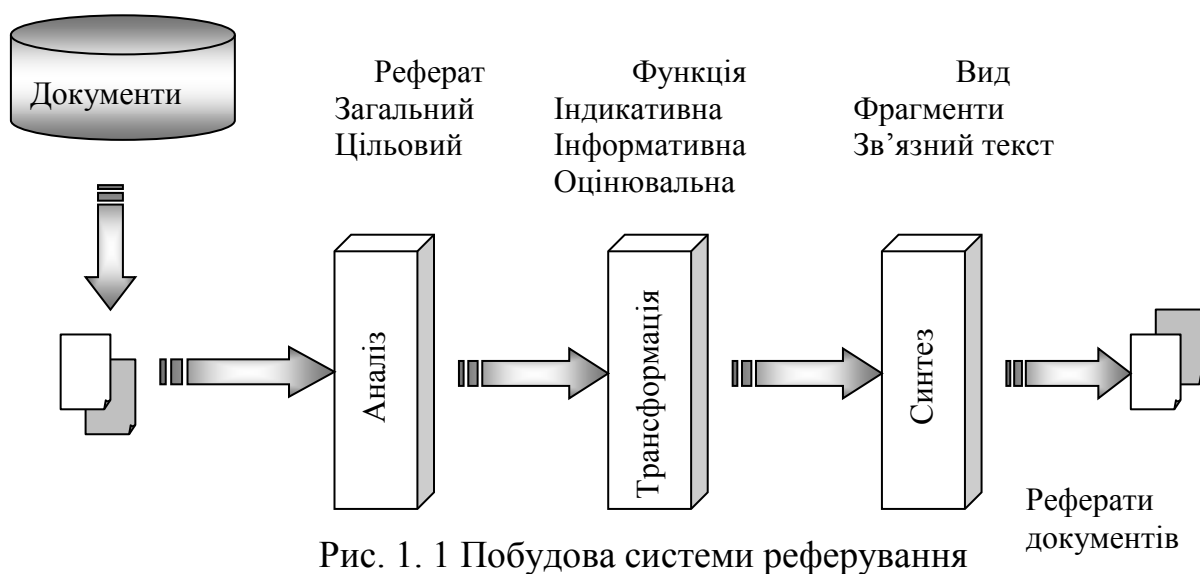


Рис. 1. 1 Побудова системи реферування

Аналіз тексту починається з вилучення із вихідного тексту лексичних одиниць (слів і словосполук), їх зважування за визначеними критеріями і виявлення масиву вагоміших. Спочатку вилучаються з вихідного тексту всі лексичні одиниці, і з них будується послідовний словниковий масив. Кожній лексичній одиниці надається попередній коефіцієнт, який залежить від її розташування у вихідному тексті. Потім вони нормалізуються за допомогою

засобів автоматичного морфологічного аналізу. Мета морфологічного аналізу полягає у зведенні слів до канонічного вигляду, тобто вилучення основ слів (словоформ з відсіченими закінченнями), а також при необхідності залучення синонімічних ланцюжків для окремих слів. Для виконання подальшого семантичного аналізу (трансформація і синтез) до кожної словоформи приєднується значення граматичних категорій (рід, відмінок, число).

Підсумовуючи все викладене вище, можна виділити два головних напрямки, за якими ведуться роботи в галузі створення засобів автоматичного укладання рефератів:

- 1) використання статистичних характеристик текстових повідомлень;
- 2) використання даних щодо граматичної та лексичної структури тексту.

1. 2. Методи автоматичного реферування

Різноманітність чинних методів реферування суттєво ускладнює появу адекватних їх класифікацій. Труднощі формулювання чітких визначень посилюються прагненням авторів до комбінованих класифікацій, намаганнями поєднати переваги різних підходів.

В. С. Пономаренко всі методи проектування інтелектуальних систем розподіляє за такими ознаками [149, ст. 32]:

- виконання технологічного процесу проектування (методи аналізу, синтезу, декомпозиції, формалізації та моделювання);
- ступінь автоматизації проектних робіт (оригінальне, типове, автоматизоване проектування);
- організація процесів проектування (різні організаційні методи).

Слід зазначити, що на відміну від інших проблем машинного опрацювання інформації, завдання автоматичного реферування, пошуки його практичних розв'язань мало привертало увагу спеціалістів. При цьому основні зусилля зосередилися на розробці методів реферування, які базуються на вилученні з вихідного тексту окремих фрагментів. Тільки останнім часом посилюється інтерес до розробки методів, пов'язаних з деякими елементами логічного опрацювання інформації.

Методи автоматичного укладання рефератів документів ґрунтуються на залученні з вихідного тексту тієї інформації, яка може бути отримана за допомогою комп'ютерних процедур.

У 70-х роках ХХ ст. запропоновано низку методик реферування, ґрунтованих на змістовому аналізі тексту першоджерел, зокрема: *поаспектне* (дозволяє виокремити більш важливі сторони й оцінити точність і повноту реферування засобом зіставлення формулювання мети дослідження та його результатів), *фасетне* (аналізує зміст визначеної сфери наукової галузі й виокремлює сітку фасетів, що відповідають числу категорій, які розкривають специфіку наукової галузі в різних її аспектах, у результаті чого укладається табличний документ) і реферування на основі *логіко-семантичного аналізу*

(вилучаються змістові опорні пункти, що несуть максимум інформації, після чого використовується змістове групування матеріалу відповідно до тематичного спрямування) [33; 71, ст. 26–28; 35; 40; 157; 175].

Аналіз літератури свідчить, що розвиток методів автоматичного реферування провадився традиційним шляхом – через удосконалення принципів і методів ручного (інтелектуального) реферування. Автоматизація, безумовно, суттєво сприяє прискоренню цього процесу, скорочує чисельність працівників, усуває будь-які спроби суб'єктивного викладення змісту першоджерел тощо, але невисока якість самих рефератів зумовлює недосконалість наявних методів автоматичного реферування.

Як справедливо зазначає А. Є. Конюшко, оптимізація реферування має базуватися на універсалізації цього процесу, запровадженні єдиних засад до семантичного згортання інформації [89].

При реферуванні слід пам'ятати, що зміст реферату є похідним від змісту документу, який реферується [251]. Звідси – необхідність установлення змістової подібності або семантичної тотожності реферату і першоджерела. Для цього залучають критерій семантичної адекватності та критерій семантичної еквівалентності. Перший застосовується для оцінки точності реферування, другий – для оцінки рівня повноти відображення змісту первинного документа в рефераті.

Наукові роботи дають вичерпний огляд спроб автоматизації реферування, формулюють наявні проблеми в цій галузі [197; 107; 136; 18]. Аналітичний огляд теоретичних і прикладних досліджень лінгвістичних і структурних характеристик реферату, їх реалізацію в автоматизованому опрацюванні містить монографія В. І. Горькової та Є. А. Борохова [55].

Аналіз літератури свідчить, що розбудова досліджень з АР поділяється на три основні періоди:

- 1) становлення основних методів АР;
- 2) розробка й застосування з метою АР методів синтаксичного аналізу;
- 3) розробка принципів АР на базі використання мов розуміння смислу.

Серед методів реферування розрізняють: статистичні, позиційні, дескрипторні, анкетні, методи, засновані на статистичних зв'язках елементів тексту, семантичні, синтаксичні, методи, засновані на дослідженні структури зв'язного тексту [15, ст. 19–23].

І. П. Севбо на додаток до статистичних і синтаксичних методів пропонує підхід, заснований на застосуванні інформаційно-пошукових мов [165].

В. П. Леонов розподіляє сукупність методів на статистичні, логіко-семантичні та лінгвістичні [108]. С. І. Гіндін – на статистичні (або частотні), індикаторні й позиційні [46, ст. 35–82].

За Х. П. Луном основний зміст реферованого джерела можна розкрити як перелік речень, найбільш значущих для даного документа. Значущими вважаються ті речення, що містять у своєму складі «скупчення» найбільш суттєвих для даного документа слів. Значущість слів обумовлюється частотою їх вживання в тексті [259, ст. 15–21]. Переважна частина сучасних програм з АР ґрунтується саме на статистичних методах.

При позиційному або аспектному підході для ідентифікації найбільш значущих речень удаються до розташування речень у тексті. Існує думка, що основний зміст первинного документа відбивається в рефераті, який складається з перших речень усіх абзаців [112, ст.89–94], або з першого, другого та останнього речень тексту [114, ст.190–197]. Позиційний метод сприяє уніфікації реферату й залучається до сучасних АР у поєднанні зі статистичним підходом. Індекссування реферату, побудованого за позиційним методом, легше за все піддається автоматизації.

За змістові одиниці тексту різними авторами пропонуються: речення, абзац, складна синтаксична, надфразова єдність тощо. У трудах М. І. Жинкіна, А. І. Новікова, В. Ш. Рубашкіна викладається денотативний, а в роботі Т. М. Дрідзе інформативно-цільовий підходи до аналізу тексту [65, ст. 34–44; 69; 132; 160, ст. 58–73]. Головний висновок щодо відповідності завданню змістового аналізу полягає в незаперечних перевагах процедури членування

тексту на окремі смислові аспекти (контексти), які характеризують тему документа.

Важливим, на наш погляд, теоретичним внеском є алгоритм аналізу початкового тексту, запропонований Г. С. Ждановою [68, ст. 12–16]. В основу методу покладено імітацію інтелектуального процесу реферування, коли з тексту вибирається інформація за ступенем її важливості.

Певну зацікавленість представляє ідея Е. Ф. Скороходько щодо вибору оптимальної процедури реферування залежно від типу семантичної структури тексту й числових характеристик семантичних відношень у цьому тексті [169, ст. 23–30; 170].

У роботі Н. І. Гендіної як один із підходів до укладання рефератів досліджуються можливості формально-змістового аналізу текстів первинних документів [45]. Відмінність цього підходу від інших більш відомих полягає в урахуванні формальних текстових ознак – маркерів (стійких словесних зворотів, що характеризують конкретні аспекти змісту).

Для розкриття семантичної структури тексту Г. Г. Белоноговим та Р. Г. Котовим запропоновано перекласти документ з природної мови певною штучною інформаційною мовою, спеціально призначеною для формального передавання змісту [9, ст. 114]. Перешкодами на цьому шляху, втім, є не складності створення такої мови, а розробка методів автоматичного перекладу текстових повідомлень з природної мови на інформаційну.

Завдання автоматичного реферування, на думку В. П. Ісакова, Г. Г. Белоногова та Б. А. Кузнєцова, зводиться до отримання на основі вихідного тексту певного «вторинного» тексту, меншого за обсягом, але такого, що відбиває головний зміст оригіналу [76, ст. 437; 11, ст. 112–113]. Сьогодні розв'язання цього завдання йде двома шляхами – квазіреферування й короткий виклад змісту первинних документів: у першому випадку формується набір фрагментів, у другому – надається короткий виклад документа.

Перший напрямок простіший і цілком піддається машинній обробці. Екстрагування робить акценти на виокремленні характерних, найбільш інформативних фрагментів методом зіставлення фразових шаблонів, у результаті чого виділяються блоки лексичної і статистичної релевантності, з яких і формуються квазіреферати [189; 275, ст. 399–408; 285, ст. 409–445].

У цьому випадку застосовується лінійна модель вагових коефіцієнтів для кожного блоку тексту відповідно до певних характеристик, а саме: розташування блоку в оригіналі, частота появи в тексті, частота використання в ключових реченнях [189]. Автоматичне визначення частотності використання окремих слів і сполучень у вихідному документі дозволяє визначати абзаци й речення, в яких тематика документа представлена найбільш точно (рис. 1.2).



Рис. 1. 2 Узагальнена схема реферування без опори на знання

Квазіреферування має ту особливість, що ґрунтується на аналізі поверхнево-синтаксичних відношень лексичних одиниць у тексті, виражених у ньому без звернення до семантичних процесів, вивчення яких поки ще недостатнє для опису властивостей будь-якого тексту. Звідси основним недоліком є те, що виділення речень, без урахування відношень між ними, призводить до формування нескладних рефератів. Деякі речення можуть бути пропущені або в них можуть зустрічатися слова чи фрази, які неможливо зрозуміти без попереднього тексту. Спроби вирішити цю проблему зводяться головним чином до виключення таких речень із рефератів або робляться посилання, які ґрунтуються на методах лінгвістичного аналізу. В ряді автоматизованих підходів створюються спеціальні інтерфейси, що допомагають визначити наявність значеннєвого розриву або «завислого»

слова. Очевидно, що такий підхід не придатний для масової обробки текстів. Якість реферування також багато в чому залежить від жанру оброблюваного тексту. Так, для великих текстів, монографій або інтерв'ю побудова якісного реферату із фрагментів вихідного документа без урахування семантичних закономірностей фактично неможлива.

У рамках квазіреферування виділяють три основних напрямки, які застосовуються сукупно в сучасних системах:

– статистичні методи, засновані на оцінці інформативності різних елементів тексту за частотою використання, що служить основним критерієм інформативності слів, речень або фраз;

– позиційні методи, що спираються на припущення про те, що інформативність елемента тексту залежить від його позиції в документі;

– індикаторні методи ґрунтуються на оцінці елементів тексту з огляду на наявність у них спеціальних слів і словосполучень – так званих маркерів важливості, як зазначає Д. П. Блюменау («на закінчення...», «було відзначено, що...», «у даній статті...», «відповідно до результатів аналізу...»), що характеризують їх значеннєву вагу [23, ст. 16–20].

Другий шлях, більш складний і поки розглядається як перспективний, тому, що припускає проведення семантичного аналізу тексту за допомогою методів штучного інтелекту й спеціальних інформаційних мов. У результаті виділяються основні аспекти змісту документа й подаються у вигляді нових фраз і речень, яких не було в тексті оригінального документа. Реалізація методу короткого викладу вимагає логічного перероблення інформації і породження речень, що передають певний зміст. При цьому необхідно знати семантичну структуру природної мови і процес мислення в його співвіднесеності з досвідом людини та системою накопичених знань обробки природних мов.

На рис. 1.3 зображено метод формування короткого викладу. Він припускає два основних підходи, що базуються на:

1) традиційному лінгвістичному методі синтаксичного аналізу речень: будується дерево, після чого скорочуються його гілки на підставі певних структурних критеріїв (дужки, вбудовані умовні або підрядні речення);

2) розумінні природної мови: дерева в цьому випадку не породжуються, а формуються концептуальні репрезентативні структури (формули логіки предикатів, семантична мережа або набір фреймів) усієї вихідної інформації, які акумулюються в текстовій базі знань [189].



Рис. 1. 3 Два основних підходи методу формування короткого викладу

Застосування баз знань може одержати більшого поширення в тих предметних галузях, для яких існують досить великі за розміром граматики і джерела знань. Важливо враховувати, що для роботи з цими джерелами або потрібні фахівці, які володіють широкими знаннями у своїй галузі, або необхідно використовувати методи машинного навчання. Для того щоб уможливити емпіричну оцінку автоматично сформованих рефератів, необхідні додаткові набори текстів та їхніх рефератів.

Для обох підходів методу формування короткого викладу завдання автоматичного реферування полягає в презентації тексту у вигляді його формального опису (після чого відбувається вибір ключового, головного в тексті) і генерації правильно побудованого тексту природною мовою.

Сьогодні розробники засобів автореферування все більше схиляються до гібридних систем, успішно поєднуючи статистичні методи й методи штучного інтелекту з застосуванням баз знань і тезаурусів. У найбільш розвинених системах реферування враховується залежність речень одне від одного, підбираються групи взаємозалежних речень, фіксуються семантичні зв'язки в

тексті і встановлюється інформативність текстових одиниць, що забезпечує зв'язність результуючих рефератів.

До таких методів належать:

Методи симетричного реферування, описані у трудах В. С. Ступіна та В. Ятско, ураховують зв'язки між реченнями на основі словника термінів предметної галузі: важливими вважаються речення – концентратори зв'язків [179, ст. 579–591; 248, ст. 33–46].

Підхід на основі розмаїтості застосовується Т. Номото й І. Метсьюмото в роботі [270, ст. 363–389]. Ідея методу полягає в тому, щоб спочатку знайти тематичні кластери документа (групи речень, що відносяться до однієї підтеми документа), після чого важливі речення виділяються з кожного кластера. Кластеризація речень виробляється за допомогою модифікованого методу k-середніх.

Концепцію фрактального реферування запропоновано С. С. Енг та Ф. Л. Венг для контенту на мобільні пристрої, які мають невеликий розмір екрану, обчислювальну потужність і низьку швидкість передачі даних [301]. Запропонований підхід використовує традиційні методи виділення речень, але тут додатково враховується інформація про ієрархічну структуру документа й необхідний «рівень абстракції» подання документа.

Таким чином, наявні засоби реферування, формуючи короткий переказ або набір фрагментів, переслідують дві основні мети: визначити основну (або найбільш важливу) думку оригіналу й прийняти рішення щодо методу скорочення (утиснення або «урізання») інформації.

Залучені методи для реферування одного документа поширюються на реферування значної кількості інших документів. Методи опрацювання припускають аналіз кожного документа й витягнення інформації в процесі перетворення і синтезу. Засоби реферування, як і раніше, виконують операції відсікання інформації, її об'єднання та узагальнення. Але простий набір рефератів кожного документа в цьому випадку не може вважатися за задовільний, тому що при наявності великої кількості рефератів вони

неминуче міститимуть надлишкову інформацію. Також засоби реферування повинні виявляти подібності й розбіжності у всіх документах шляхом порівняння й злиття презентацій змісту документа, отриманих на етапі аналізу [277, ст. 193–207; 264, ст. 35–67; 276, ст. 469–500].

Оскільки повідомлення про одні й ті самі події, які відрізняються одне від одного, з'являється в різних джерелах інформації, у публікації Дж. Кербонел та Дж. Голдстейн відображено засоби реферування, що відкидають надлишкову інформацію з усіх повідомлень і видають короткий виклад [217, ст. 335–336].

Р. К. Ендо акцентує увагу на тому, що засоби реферування також можуть відслідковувати загальну тему в різних повідомленнях і представляти результати своєї роботи за допомогою діаграм і графіків [211, ст. 79–88].

Ще однією складністю в розробці систем реферування є оцінка методів реферування, тому методи укладання й оцінювання рефератів повинні розвиватися паралельно.

Методи оцінювання рефератів визначають адекватність, вірогідність і користь реферату стосовно оригінального тексту. За Ж. Селтон, на сьогодні відомі дві методики [277, ст. 193–207]. Перша – оцінка «зсередини» (або нормативна оцінка). Користувачі оцінюють якість реферату, аналізуючи сам реферат, зв'язність тексту, роблять висновок про те, наскільки добре реферат відбиває основні ідеї оригіналу, або порівнюють його з ідеальним рефератом, написаним автором вихідного тексту або іншим фахівцем.

Другий метод – оцінка «ззовні». Користувачі оцінюють якість реферату за тим, як він діє на завершення тієї чи іншої роботи, наприклад, чи допомагає він знайти джерела інформації з даного питання або наскільки добре він дозволяє відповісти на певні питання, що ставляться до всього змісту тексту.

Зауважимо, що жодна з цих оцінок не може вважатися повністю задовільною. Так само, як існує безліч способів описати якусь подію, користувачі можуть визнати прийнятними кілька рефератів, чи то налаштовані на користувача, чи на загальні короткі виклади, або набори викладів.

Е. Ф. Скороходько пропонує такі методи оцінювання реферату: *точність і повнота* (частка істотної інформації вихідного тексту в рефераті), *зв'язність, надмірність, економічність, коефіцієнт стиснення* (розмір реферату до розміру тексту) і *швидкість реферування* (ступінь складності математичного забезпечення, необхідність спеціальних словників) [170, ст. 157–161].

Існує єдина думка про необхідність кращих методів оцінки. Однак, більшість завдань ще не вирішено, у тому числі, зберігається необхідність у досконалих методологіях створення рефератів. Проте багато технологій уже працюють, і можна чекати, що інструменти реферування будуть відігравати вирішальну роль у завоюванні широких інформаційних просторів у майбутньому.

1. 3. Огляд лінгвістичних систем автоматичного реферування

При розробці систем АР однією з ключових проблем була й залишається проблема аналізу семантичної структури вихідного тексту з метою визначення фактографічної інформації в межах тематики конкретного тексту з подальшим її стисненням і синтезом. Наявні системи реферування не вирішують цих завдань, а процедура АР зводиться частіше до побудови квазіреферату на основі статистичного й позиційного аналізу тексту як способу оцінки його інформативності для екстрагування найбільш інформативних фрагментів тексту. Водночас сьогодні пильну увагу приділяють розробці систем опрацювання текстів на семантичному рівні.

На ринку програмного забезпечення існує безліч програм автореферування, які виділяють значущі речення в тексті, використовуючи переважно статистичні алгоритми або індикатори – слова-підказки. Деякі виробники пропонують інструменти для реферування: функція AutoSummarize в Microsoft Office, Inxight Summarizer (компонент пошукового механізму AltaVista). Для реферування текстів російською мовою існує компонент системи ОРФО, що надається компанією «Информатик» (www.informatic.ru). Система дозволяє перевіряти орфографічні та граматичні помилки, синтаксис, стиль, робити квазіреферат документа, вилучати ключові слова тощо. Програма Лібрето компанії «МедіаЛінгва» (www.medialingua.ru) виділяє ключові слова, виконує автореферування російських і англійських документів. Обидві ці програми можуть бути вбудовані в середовище Microsoft Word.

На жаль, їхні можливості обмежені виділенням і вибором оригінальних фрагментів із вихідного документа та поєднанням у короткий текст. Підготовка ж короткого викладу передбачає передавання основної думки тексту, і не обов'язково тими самими словами. Наявні системи автоматичного реферування моделюють, зазвичай, не сам процес інтелектуального реферування (процеси читання, розуміння, осмислення першоджерела), а кінцевий результат – витяг із тексту фрагментів і оформлення реферату за

заданою програмою. Саме це є сутністю процесу реферування (компресії тексту), інша справа, що виконується згортання не в результаті змістового аналізу тексту, а в результаті певних формальних процедур, що мають обмежені можливості. Тому, коли ринок недорогих традиційних систем автоматичного реферування був практично заповнений, з'явилися нові гібридні рішення, побудовані на основі синтаксичних аналізаторів.

Так, Human Language Technology Sector of the Information Society Technologies (IST) Programme 1998–2000, який координує Європейський Союз, основні розробки присвячує автоматизації процесу синтаксичного аналізу для будь-яких систем АОІ, у тому числі й АР [246]. Мета цього аналізу – автоматична побудова функціонального дерева фрази, тобто пошук взаємозалежності між різнорівневими елементами речення.

Синтаксичний аналізатор Ergo Linguistic Technologies Parser, розроблений Дерекком Бікертоном і Філіпом Браліком (Університет, Гонолулу), має відому схему аналізу і предметне вираження [222]. ERGO орієнтує свій парсер на використання інтерфейсів у вигляді питань і відповідей. Це поки що єдина компанія, яка має парсер, здатний визначати тип запитання (питання до підмета, прямого або непрямого додатка чи обставини) і швидко конструювати відповідь.

Вдалих синтаксичний аналізатор Functional Dependency Grammar створено дослідниками Гельсинського університету, які пізніше заснували фірми Lingsoft і Conexor [229]. Рання версія під назвою «ENGCG» (English Constraint Grammar) була використана для реферування найбільшого у світі корпусу – Bank of English, що належить видавництву Collins/Harper Publishers. Особливістю даного аналізатора є те, що в разі неможливості зняття багатозначності, він видає декілька варіантів аналізу або не добудовує дерево для даної частини пропозиції.

Оригінальний підхід до синтаксичного аналізу тексту – Link Parser започатковано в Carnegie-Melon University [258]. Це єдиний синтаксичний аналізатор з опублікованими он-лайн кодами. Якщо більшість систем

синтаксичного аналізу використовують структури рівня іменникових і дієслівних груп у побудові дерева фрази, то Link Grammar, покладена в основу Link Parser, використовує інформацію про типи зв'язків.

Найбільш цікаві є передові, в експериментальному плані, системи, що враховують зміст тексту, але поки, на жаль, не вирішують завдання розуміння тексту в повному обсязі.

До систем, які враховують змістові аспекти тексту, належать система АДАМ (ADAM – Automatic Document Abstracting Method), розроблена на початку 1970-х рр. у реферативній службі Chemical Abstracts Service (США). Вона розрахована на екстрагування речень, що містять змістові аспекти з інформацією про мету й результати роботи. В екстракт не включаються відомості експериментального й методичного характеру, а також нетекстова інформація (таблиці, графіки, схеми, рівняння, формули тощо). Екстрагування в даній системі робиться на основі словника індикаторів – списку контрольних слів. Семантична вага речення встановлюється відповідно до концепції авторів системи – формувати екстракт на основі методу відмови від речення (method for rejecting sentences). Це означає, що речення, в якому висловлюється суб'єктивна точка зору автора, загальні міркування або заперечення, висновки та інша інформація, виключається з тексту.

Певну схожість із системою АДАМ має система «АНАСИН», розроблена С. А. Карасьовим, яка включає аналіз і синтез тексту. Програма вилучає з першоджерела ключові фрази і класифікує за ступенем важливості, а також усуває штучні речення, що містять посилання на літературу, табличний і графічний матеріал, неінформативні фрази. Аналіз тексту здійснюється за допомогою словника, а класифікація речень за ступенем важливості робиться за допомогою семантичних характеристик [79, ст. 22–28].

Система SUSY (Summarizing System) розроблена в Удинському університеті (Італія) і складається з двох компонентів, які працюють послідовно: перший взаємодіє зі споживачем, другий опрацьовує текст. Процедура аналізу здійснюється програмою «Аналізатор», яка використовує

інформацію про слова, конструкції, що містяться в словнику, і набір семантичних правил. Ці правила визначають властивості й відношення між елементарними поняттями, важливими з точки зору предметної галузі.

Технологія SemLP (Semantic Language Processing) пропонує весь спектр рішень, що вимагає опрацювання текстів природною мовою: пошук документів у локальних і глобальних мережах, сортування й класифікацію документів, автоматичне анотування й реферування, перевірку коректності текстів, питально-відповідні та діалогові системи, машинний переклад, навчальні програми, системи підтримки розробки баз знань, тезаурусів і онтологій [www.semlp.com]. Щодо реферування, то програма дозволяє реалізувати в обмеженому вигляді автоматичний синтез коротких (у кілька слів) простих фраз або речень. У цілому завдання реферування включає визначення тематики документів, виділення ключових (за темами) слів і фраз із урахуванням змісту, пошук речень, що містять ключові слова й фрази, та синтез на цій основі фраз і речень, що відбивають основні теми тексту.

Особливим є «обчислювальний» та «функціональний» підходи до мови. Текст розглядається не як статична структура, елементи якої потрібно розпізнати, а потім якимось чином перетворювати, а як запис програми або алгоритму. Слова подаються функціями, які породжують нові конструкції. При такому підході аналіз тексту перетворюється в згортання речень із їхньої вихідної форми в єдину суперпозицію функцій.

В основі технології міститься кілька базових елементів:

- 1) словник з описом моделей керування слів і їхньої семантики;
- 2) система семантичних класів;
- 3) система типів, що описують аргументи слів (семантико-граматичні типи);
- 4) аналізатор, який будує формальне подання тексту.

Поточна реалізація аналізатора тексту забезпечує деталізацію семантичних описів, достатню для створення визначень у базі знань і ситуацій для тезаурусів і онтологій.

Одним із перших відомих комерційних проєктів автореферування була система Inxight Summarizer, створена в 1995 році (Дослідницький центр корпорації Ксерокс, Пало Альту). Вона паралельно задіює кілька відомих алгоритмів реферування й оцінки якості рефератів. Крім того, система розповсюджувалася не тільки у вигляді готової програмної цілісності, але й у вигляді модулів реферування у складі бібліотек для платформ Win32 і Solaris [249].

Компанія British Telecommunications Laboratories для експериментальної онлайн-платформи TranSend свого часу створила Prosum – систему реферування, реалізовану у вигляді cgi-сценарію, яка вбудовується в сторінки Web-сайтів [273].

Водночас існують системи, які намагаються вирішити завдання розуміння тексту через відмову від повного розуміння всіх аспектів змісту: потрібний інформаційний аспект задається або мовою системи (дескрипторна мова), або зовнішньою граматиною.

У системі ФРАП започатковано спробу створити модель розуміння тексту, в рамках якої можливо було б здійснити систему автоматичної побудови реферату. Перекладу підлягає лише опанована системою частина змісту вихідного природного тексту. Оскільки будь-яке автоматичне розуміння відбувається з утратою частини інформації, важливо, щоб ці втрати були змістовно контрольовані, тобто охоплювали найбільш периферійні частини тексту. «Розуміння» складається з виявлення головного, домінантного аспекту змісту й перебудови його в ті одиниці, якими володіє сприймаюча система-реципієнт (наприклад, одиниці тезаурусу термінів, схеми предметної галузі тощо) [107].

Зміст цього підходу полягає не у вимозі завжди замінити переклад його рефератом, а в створенні реферату як визначальника домінантного аспекту. Застосовується реферування для передавання основного змісту вихідного тексту іншою мовою, тобто реферування та переклад.

У роботі Р. Брендон описано систему ANES, в якій важливі слова виділяються на основі підходу TF*IDF [215, ст. 675-685]. Мабуть, єдина відмінність від традиційного підходу – виявлення анафоричних зворотів, які дозволяють виключити з реферату речення з «завислими» відсиланнями до попередніх речень.

Одна з перших промислових програм автореферування з елементами семантичного аналізу – Extractor – була створена в Інституті інформаційних технологій Національної дослідницької Ради Канади [223]. Вона забезпечує виділення із вхідного/вихідного тексту найбільш інформативних іменникових груп. Основний модуль Extractor застосовується в програмних продуктах фірм ThinkTank Technologies і Tetranet, а також у ряді пошукових систем.

Технологія аналізу текстової інформації WebFountain, розроблена в дослідницькому центрі IBM Almaden Research Center, аналізує слабкоструктуровані й неструктуровані дані, які одержують з Інтернету [231]. Новації полягають у технологіях контент-аналізу й структурування інформації, які забезпечують точний тематичний пошук.

Компанія «МедіаЛінгва» розробила систему «Анотатор SDK 1.0» [www.medialingua.ru/annotator.html], що є набором засобів, призначених для автоматичного реферування документів російською та англійською мовами. Нею для кожного речення вхідного тексту на основі ймовірнісних моделей, доповнених лінгвістичними методами й словниками, обчислюються вагомі коефіцієнти. З найбільш значущих і незалежних речень складається реферат розміром, що може бути налаштований користувачем. Для надання реферату більшої зв'язності вихідні речення можуть бути переформульовані.

Свого часу російська компанія «Мікросистеми» (MicroSystems) випустила програму TextReferent, яка надає можливість укладати квазіреферати й дозволяє налаштовувати «подробиці» реферату. Систему TextAnalyst було розроблено цією же компанією як інтегрований інструмент у текстовий редактор Microsoft Word 97 і Microsoft Word 2000 для аналізу змісту текстів, пошуку інформації і формування електронних архівів [286]. Якщо

TextReferent опановує лише одну з функцій системи TextAnalyst – автоматичне укладання реферату, – повна версія системи надає користувачеві можливості виділення іменникових груп і побудови на їхній основі індексу у вигляді семантичної мережі. Матеріал текстів розподіляється за тематичними класами. В результаті аналізу змісту тексту з автоматичним формуванням семантичної мережі з гіперпосиланнями виходить значеннєвий портрет тексту в термінах основних понять і їхніх значеннєвих зв'язків у вигляді ієрархії тем і підтем. Інформація ранжується за ступенем вагомості з можливістю варіювання деталізації її дослідження. Текст автоматично індексується і перетворюється в гіпертекст. Програма не залежить від предметної галузі, призначається для автоматичного аналізу неструктурованої текстової інформації і формує гіпертекстову структуру, в якій семантична мережа дозволяє здійснити ефективну навігацію по тексту. Втім тези не зв'язуються між собою стилістично, а беруться з тексту й розташовуються в порядку їх появи в тексті. Речення реферату забезпечені відсиланнями до відповідних місць вихідних текстів, що дозволяє переглянути більш детально контекст тези, який цікавить. Подробиці реферату налаштовуються при зміні кількості речень, що формують його.

Обидві програми TextAnalyst і TextReferent сьогодні знайшли свій розвиток у потужній системі PolyAnalyst компанії Megaputer. Пакет PolyAnalyst використовує методи Text Mining для автоматичного аналізу чисельних і текстових даних.

Системи класу Text Mining, які включають елементи автореферування, розроблені IBM (Intelligent Miner), Silicon Graphics (SGI Miner), Integral Solutions (Clementine), SAS Institute (SAS), Oracle (Oracle Text).

Порівняно з традиційними підходами, використання технологій Text Mining при аналізі ресурсів Інтернет уже сьогодні забезпечує, поряд із включенням робочих місць користувачів у динамічний інформаційний простір, одержання оперативних кількісних і якісних аналітичних зрізів, що раніше було практично неможливим. Крім того, зростання обсягів мультимедійної

інформації в мережі Інтернет робить його також дуже важливим об'єктом для обробки засобами реферування. Технології автореферування повинні опрацьовувати дані різного типу на етапах аналізу й синтезу, реалізуючи інтеграцію інформації всіх типів.

Запропоновано перші прототипи для реферування великої кількості документів і підготовки рефератів для мультимедійних джерел. Робота над засобами реферування різномовних джерел тільки-но починається.

Прикладом багатофункціональності сучасних засобів реферування є автоматичне формування *дайджестів* [55; 105, ст. 199]. Останній є реферованим текстом або джерелом гіперпосилань на основі аналізу документів, що відбивають тенденції всього вхідного інформаційного потоку (рис.1.4). При укладанні дайджестів методи автореферування одного документа поширюються на масив із великої кількості документів.



18 (200790)	11-MAY-2000	CNN Today
		
Length: 00:00:21	THE CASE OF ELIAN GONZALEZ WAS HEARD TODAY AT THE U.S. COURT OF APPEALS FOR THE 11th CIRCUIT IN ATLANTA.	
Real Video	128K	
Similar Stories	BNN Stories	
PERSON	ELIAN GONZALEZ	JUAN MIGUEL GONZALEZ
ORGANIZATION	U.S. COURT OF APPEALS	
LOCATION	ATLANTA	FLORIDA

Рис. 1. 4 Реферат як засіб пошуку, перегляду й реферування телевізійних новин

Такий мультимедійний реферат, виданий на запит пошукового механізму, містить у собі ключові речення, а також інформацію про важливих персон, організації та місце розташування, що супроводжуються відео. Натискання на кадри відео активізує відтворення відеофрагменту за допомогою відповідного плеєра. Система Broadcast News Navigator пропонує також посилання на інформаційні повідомлення, які стосуються запиту користувача [267, ст. 381–391].

Справитися з завданням створення коротких, але змістовно повноцінних аналогів текстів – рефератів може лише система, здатна аналізувати зміст текстових документів не за формальними, а за змістовими ознаками.

Наявні системи автоматичного реферування засновані на статистичних і позиційних методах автоматичного опрацювання інформації. В цих системах не йде мова про повноцінний семантичний аналіз інформації, всі вони вирішують лише завдання автоматичного екстрагування найбільш значущих речень.

Створення ж високоякісного автоматичного реферату потребує проведення аналізу логіко-сислової структури текстів, узагальнення смислу й наявності в системі для їх здійснення розвинутих словникових засобів і онтологій. Саме цьому присвячено запропонований у даному дослідженні підхід до моделювання процесу узагальнення. Метою запропонованого підходу є побудова моделі реферату як набору типових для індикативних рефератів синтактико-семантичних конструкцій (структур) із загальним метазначенням з подальшим змістовим конструюванням реферату, яке передбачає наповнення синтаксем поняттями відповідних класів з онтологій предметних галузей.

Таким чином, наступним кроком у дослідженні узагальнення в процесі реферування для удосконалення системи АР є моделювання з точки зору саме узагальнення, для чого потрібне створення відповідних онтологій.

1. 4. Моделювання знань у системах автоматичного реферування

Моделювання – один з найбільш розповсюджених способів вивчення різних процесів і явищ [29, ст. 42]. Звичайно розрізняють фізичне й математичне моделювання. При фізичному моделюванні, що має обмежену сферу застосування, модель відтворює досліджуваний процес (оригінал) зі збереженням його фізичної природи. Ширші можливості у математичного моделювання, під яким розуміється спосіб дослідження різних процесів шляхом вивчення явищ, що мають різний фізичний зміст, але описуються однаковими математичними співвідношеннями.

Формалізації будь-якого реального процесу передують вивчення структури складових його явищ. Результатом цього є змістовний опис процесу реферування – спроба чітко викласти закономірності і поставити прикладне завдання. Змістовний опис, складений у результаті ретельного вивчення процесу реферування, є вихідним матеріалом для наступних етапів формалізації: побудови формалізованої схеми й математичної моделі.

Формалізована схема процесу є проміжною ланкою між змістовним описом і математичною моделлю. Однак, на думку Н. П. Бусленка, вона розробляється не завжди, а лише тоді, коли через складність досліджуваного процесу або труднощі формалізації деяких його елементів безпосередній перехід від змістовного опису до математичної моделі виявляється неможливим або недоцільним [29, ст. 45].

Математична модель, що описує формалізований процес функціонування системи, здатна охопити тільки основні, притаманні йому властивості, залишаючи осторонь несуттєві, другорядні чинники.

Складна система є багаторівневою конструкцією із взаємодіючих елементів, поєднаних у підсистеми різних рівнів. Відповідно, математична модель складної системи складається з математичних моделей елементів і математичної моделі взаємодії між елементами.

Сьогодні моделювання, що розглядається як сукупність методик і технічних засобів, побудованих на підставі тих чи інших природнонаукових

законів, зазнало широкого розвитку [191, ст. 5]. Цифрове моделювання в ряді галузей науки і техніки стало найпершим методом досліджень. Щонайбільше ефективно його застосовування при вивченні випадкових процесів.

Моделювання передбачає побудову діючої математичної або реальної моделі з іманентними властивостями, адекватними до властивостей або співвідношень природної або технічної системи. Таким чином, можна імітувати роботу системи і приймати попередні рішення щодо оптимізації її характеристик, фактично не маючи діючого пристрою [191, ст. 21].

У цьому разі спостереженню і дослідженню піддається не сам об'єкт, а його аналог (модель), що використовується при описуванні лінгвістичних явищ у межах досліджень з прикладної лінгвістики, оскільки «...спілкування людини з комп'ютером може здійснюватися лише за умови, що в пам'ять машини уведена певна модель, яка становить скорочений опис природної мови» [148].

Дослідження методом моделювання починається з розроблення спеціальної моделі й триває в ході роботи з нею [4].

Моделі використовуються для вивчення і опису внутрішньої побудови оригіналу (структурні моделі), його поведінки (функціональні моделі) та побудови й поведінки оригіналу одночасно (структурно-функціональні моделі).

Зазначимо головні положення, які відносяться до гносеологічних аспектів моделювання. Модель має:

- виступати особливим засобом відображення дійсності, який уможливлено на основі аналогії;
- бути не складніша за сам оригінал;
- бути вільною від протиріч (логічно коректною), вичерпною і максимально простою;
- загальний характер, що дозволяє застосовувати її для опису різних натуральних об'єктів;

– володіти пояснювальною силою (експланаторністю), суть якої полягає в здатності моделі розкривати і пояснювати невідомі дотепер властивості оригіналу;

– містити евристичні можливості, тобто давати такі знання, які самі стають джерелами нових ідей і теорій;

– передавати суть оригіналу, тобто відбивати і відтворювати ті реальні та істотні його риси, які є найбільш важливими для даного експерименту. При цьому побічні, не релевантні з погляду досвіду, властивості можуть не враховуватися [146].

Побудова моделі передбачає [4]:

- 1) фіксування фактів, що вимагають пояснення;
- 2) висування гіпотез для пояснення фактів;
- 3) реалізацію гіпотез у вигляді моделей, які не тільки пояснюють вихідні факти, але й передбачають нові, невідомі факти;
- 4) експериментальну перевірку.

Структура моделі повинна бути добре відомою досліднику, щодо оригіналу – необхідно мінімум даних, які свідчать, що елементи та їхні відношення в оригіналі та в моделі перебувають у відношенні повної або часткової подібності. На основі цих даних будується процедура перенесення на оригінал тієї інформації, що була отримана в результаті спостережень над структурою і роботою моделі.

Сучасний етап у розвитку систем опрацювання інформації взагалі характеризується як етап формалізації і застосування знань. На сьогодні в цілому сформувалася і активно розбудовується нова науково-технічна дисципліна – інженерія знань, яка об'єднала зусилля фахівців у галузі прикладної інформатики, математики, лінгвістики, психосемантики. Наріжним напрямком інженерії знань, на думку В. Ш. Рубашкіна, є пошук мови логічної систематизації знань – важливої ланки між мовами опису знань, що застосовуються в сучасних системах штучного інтелекту, і знаннями фахівця [161].

Задати єдиний алгоритм моделювання процесу розуміння текстової інформації неможливо. Кожен з розроблених алгоритмів висвітлює лише певні аспекти знань і в прикладному плані орієнтований на визначений клас завдань [74, ст. 2–4].

Формалізований аналіз загального стану справ із штучного інтелекту висвітлює безперечно зацікавленість до засобів представлення знань [150]. Усі намагання автоматизації реферування, запроваджені останнім часом, дійшли однозначного висновку – вирішення цього завдання потребує відповідних засобів подання знань. Сьогодні в усьому світі ведуться дослідження зі створення концептуальних моделей подання знань і їх залучення для опрацювання інформації в різних сферах діяльності [102; 283, ст. 161–197].

За визнанням багатьох авторів [84, ст. 20–27; 283], термін «подання знань» – це спосіб кодування знань у базі знань або формальна система, що застосовується для формалізації знань. З цим терміном пов'язується певний етап у розвитку програмного забезпечення. Послідовний розвиток структур даних обумовив їхню якісну зміну й перехід від подання даних до подання знань.

Подання знань базується на аналізі текстів, що відносяться до певної предметної галузі:

- 1) виділення й експліцитне подання структури змісту окремого тексту;
- 2) об'єднання окремих структур змісту конкретних текстів і побудова на цій основі загальної системи знання відповідної предметної галузі [116].

Практика розробки систем, заснованих на знаннях для складних предметних галузей і завдань засвідчила, що в кожній предметній галузі існує деяка структура, яка займає проміжну ланку між поданням знань і моделлю предметної галузі (базою знань), тобто між структурою, яка визначає, що може бути представлено, і тим, що існує в дійсності. З 80-х років розпочато систематичне вивчення даної структури, а в наукових трудах з'явилася назва «онтологія предметної галузі» [193; 156; 295].

Поширення мережі Інтернет призвело до чисельності розрізаних джерел і відомостей (баз даних і знань, окремих інформаційних ресурсів), що характеризуються різними засобами подання, форматами й мовами опису. Щоб звести різноманітно представлену інформацію до загальнозрозумілого вигляду і в такий спосіб організувати її спільне й багаторазове використання, знадобилися методи стандартизації подання інформації.

Сьогодні у сфері інформаційного менеджменту простежується перехід від керування *даними*, що характеризують кількісну сторону інформаційних процесів, до керування *знаннями* як якісної складової цих процесів. Якщо дані є набором розрізаних фактів, представлених у різній формі, то знання – це безліч відношень (обмежень, функцій, правил) у семантичній моделі певної проблемної галузі, які так само, як і дані, відображаються у вигляді формул, тексту, файлів, інформаційних масивів.

Т. А. Гаврилова і В. Ф. Хорошевський визначають дані як «окремі факти, що характеризують об'єкти, процеси і явища предметної галузі, а також їхні властивості», а знання – «закономірності предметної галузі (принципи, зв'язки, закони), отримані в результаті практичної діяльності і професійного досвіду, а також допомагають фахівцям при вирішенні завдань у цій галузі» [41, ст. 19].

Існують різні класифікації видів знань, але в кожній з них, як правило, виділяють *процедурні* й *декларативні*, *поверхневі* і *глибинні* знання [41, ст. 19–21].

Перші системи штучного інтелекту працювали з процедурними знаннями, які керували даними. В міру вдосконалювання систем штучного інтелекту пріоритет даних поступово змінювався і все більша частина задіяної в них інформації зосереджувалася в структурних даних (таблицях, списках, абстрактних типах даних), тобто збільшувалася роль декларативних знань, записаних на мовах подання знань, наближених до природної і зрозумілої неспеціалістам.

Поверхневі знання – це знання про видимі взаємозв'язки між окремими подіями й фактами в предметній галузі, а глибинні – абстракції, аналогії, схеми, що відображають структуру й природу процесів у предметній галузі. Ці знання пояснюють явища і можуть залучатися до прогнозування поведінки об'єктів.

До числа найбільш відомих моделей подання знань відносяться *продукційні, мережні, фреймові, логічні й нейроподібні моделі* [41; 66; 283]. Кожна з них має свої сильні і слабкі сторони.

Продукційна модель або модель, заснована на імплікативних правилах, дозволяє подати знання у вигляді речень типа «Якщо (умова), то (дія)». Найчастіше такий вид моделі застосовують у промислових експертних системах, тому що вона приваблює розроблювачів своєю наочністю, високою модульністю, легкістю внесення доповнень і змін, а також простотою механізму логічного висновку. В моделях цього типа використовуються деякі елементи логічних і мережних моделей. З логічних моделей запозичена ідея правил висновку, які тут називаються продукціями, а з мережних моделей – опис знань у вигляді семантичної мережі. Замість логічного висновку в продукційних моделях з'являється висновок з опертям на знання. Однак процес висновку в цих системах менш ефективний, ніж інші, тому що потребує більше часу через перевірку прийнятності правил (продукцій). Крім цього в системах продукцій відсутні засоби для встановлення ієрархії правил [66, ст. 311].

В основу мережних моделей покладено конструкцію, названу раніше семантичною мережею, яка представляє собою орієнтований граф, вершини якого – поняття, а дуги – відношення між ними [41, ст. 23]. Основною перевагою такої мережі є те, що вона краще відповідає сучасним уявленням щодо організації тривалої пам'яті людини. Е. Ф. Скороходько, з лінгвістичної точки зору переваги семантичних мереж, бачить у можливості відображати велику мовну систему в повному обсязі і з усіма деталями [170, ст. 4].

Мережеве моделювання тексту включає три основні операції: виявлення семантичних зв'язків між елементами тексту, їх подання у зручній, для подальших досліджень, формі і встановлення закономірностей, що характеризують досліджуваній об'єкт [132, ст. 101].

Семантична мережа є ідеальним засобом представлення будь-якого системно-структурного утворення і корисна при вирішенні багатьох теоретичних і практичних завдань, особливо пов'язаних із проектуванням лінгвістичного забезпечення систем штучного інтелекту. Недоліком моделі є складність організації процедури висновку.

Залежно від типів зв'язків, застосованих у моделі, розрізняють класифікаційні мережі, функціональні мережі і сценарії. У класифікаційних мережах використовують відношення структуризації. Такі мережі дозволяють у базах знань вводити різні ієрархічні відносини між інформаційними одиницями. Функціональні мережі характеризуються наявністю функціональних відносин. Їх часто називають обчислювальними моделями, тому що вони дозволяють описувати процедури «обчислень» одних інформаційних одиниць через інші. Для опису стереотипних знань (відомі стандартні ситуації реального миру) найчастіше використовують сценарії. Під сценарієм розуміється формалізований опис стандартної послідовності взаємозалежних фактів, що визначають типову ситуацію предметної галузі. В інформаційних системах сценарії використовуються в процедурах розуміння природно-мовних текстів, планування поведінки, навчання, прийняття рішень, керування змінами середовища тощо.

На противагу моделям інших типів у фреймових моделях фіксується жорстка структура інформаційних одиниць, що називається протофреймом. Під фреймом розуміється абстрактний образ для подання якогось стереотипу. Р. Шенк поділяє фрейми на фрейми – візуальні образи і фрейми – сценарії (скрипти) [200]. Перевага фрейму як моделі подання знань у тім, що він відбиває концептуальну основу організації пам'яті людини, її гнучкість і наочність. Відповідно до цього підходу знання про світ можуть бути

представлені у вигляді значної сукупності певним чином структурованих даних, що є стереотипними ситуаціями [122].

Фрейм можна уявити собі у вигляді сітки, що складається з вузлів і зв'язків між ними. «Верхні рівні» фрейму чітко визначені, оскільки утворені поняттями, які завжди справедливі щодо передбачуваної ситуації. На більш нижчих рівнях є термінали чи «осередки» (слоти), які повинні бути заповнені певними прикладами або даними.

Кожним терміналом можуть встановлюватися умови, які визначаються маркерами. Семантично близькі між собою субфрейми, фрейми і суперфрейми – ієрархічно упорядковані елементи, що утворюють системи фреймів. Більш складними умовами задаються відношення між поняттями, внесеними в різні термінальні вершини [50].

Фреймовий підхід активно використовується в діалогових системах: «...по суті, пункт діалогу – це фрейм діалогової ситуації. Сценарій діалогу при цьому формується як система фреймів діалогових ситуацій» [87, ст. 399]. Даний підхід також використовується при описуванні семантичної і синтаксичної структур текстів і при створенні онтологічних систем.

Як спосіб опису знань у науковій літературі також виділяють логічні моделі, засновані на класичному вирахуванні предикатів 1-го порядку, коли предметна галузь або завдання описується у вигляді набору аксіом. Вирахування предикатів 1-го порядку в промислових експертних системах практично не використовується. Ця логічна модель застосовна головним чином в дослідницьких системах, тому що пред'являє високі вимоги й обмеження до предметної галузі. В промислових же експертних системах застосовуються різні її модифікації і розширення.

З метою моделювання лінгвістичних процесів може бути застосована теорія мереж Петрі, що є математичним аналогом семантичної сітки у вигляді математичної моделі дискретних систем з паралельно функціонуючими і асинхронно взаємодіючими компонентами, яка є математичним інструментом для моделювання систем, а також аналізу їхньої подальшої поведінки [91;

126]. На основі інформації, отриманої в результаті аналізу, здійснюється оцінювання системи. Таке моделювання дозволяє знайти рекомендації з удосконалення, зміни й оптимізації системи.

Досить новим методом моделювання інформаційних процесів, що активно сьогодні розвивається, є моделювання нейронних мереж [245, ст. 2552–2554]. Оскільки будова й закони функціонування нейрона дуже складні, при моделюванні нейронів користуються їх спрощеними описами – нейроподібними мережами (НпС).

Розрізняють два типи НпС. У першому з них вузлами є формальні елементи, що описують окремі нейрони, в іншому – формальні елементи, що відповідають нейронним ансамблям. НпС, вузли якої відповідають нейронним ансамблям, є семантичною сіткою. Важливою властивістю НпС є здатність до самоорганізації, самоадаптації для поліпшення якості функціонування. Це досягається шляхом навчання сітки, алгоритм якої задається набором навчальних правил. Моделювання лінгвістичних процесів для систем АОТ за допомогою НпС на сьогодні не практикується, однак, на думку авторів, може стати цілком можливим вже в найближчому майбутньому.

Крім недосконалості кожної з моделей подання знань на вибір способу їхнього подання істотно впливає природа самих знань галузі, що моделюється. Більше того, при вирішенні складних завдань подання різного виду знань з самого початку припускається використання різних способів подання. В результаті база знань може мати багаторівневу структуру і тоді функціонування інтелектуальної системи здійснюється за принципом дошки оголошень [66, ст. 317].

Універсальним механізмом для опису структури знань стає *онтологія*, яка включає в себе словник, засобами якого описуються знання, і саму структуру, що описується. За допомогою словника онтологія накладає обмеження на застосовану семантику.

У даному дослідженні моделювання знань у системах АР пропонується здійснювати шляхом створення онтологій. Обраний метод подання знань –

досить новий і містить окремі елементи вищеописаних підходів фреймового й мережевого моделювання. Отримана модель знань є складовою частиною моделі процесу узагальнення змісту взагалі.

Прикладна лінгвістика має специфічну наукову проблематику, пов'язану передусім з дослідженням тексту, розробкою адекватних і практичних методів аналізу й синтезу мовних виразів, завдань морфологічного, синтаксичного, семантичного аналізу й синтезу текстів при автоматичному реферуванні, автоматичному перекладі в інформаційно-пошукових та інших спеціалізованих системах [25, ст. 3–16; 92].

Аналіз стану розробок у галузі автоматичного реферування дозволяє зробити висновок, що системи АР узагальнюють проблеми моделювання розуміння текстової інформації. Розв'язання цього завдання, враховуючи накопичений досвід у галузі лінгвістики, штучного інтелекту (ШІ), інформатики, вийшло на стадію об'єднання в єдину методологічну систему.

Ряд досліджень [10; 168, ст. 15–120; 47] показав, що при розробленні діючих моделей розуміння тексту виникають проблеми внаслідок таких чинників:

- зміст одних речень залежить від змісту інших;
- для вилучення загального змісту тексту є важливим зміст не тільки елементів речень, але і зв'язків між ними;
- для змістового запису тексту необхідний перехід від семантичного подання окремих речень до семантичного подання надфразових одиниць і потім усього тексту;
- необхідні виявлення і фіксація таких невід'ємних характеристик тексту, як тема й композиційна структура;
- значення окремого речення залежить від його контексту.

Розв'язання завдання створення семантичного компонента, що забезпечує повноцінну роботу системи автоматичного реферування, має базуватися на вивченні і систематизації глибинних механізмів зв'язного тексту. При цьому треба враховувати, що текст – це єдиний механізм, який

складається з трьох різних за своєю природою механізмів: семіотичної системи, мовної системи й системи знань про світ. Лише при такому підході з'явиться можливість перейти від «псевдозмісту» до змісту, від квазірефератів до рефератів.

Особлива увага в дослідженні тексту приділяється вивченню механізму міжфразових зв'язків як текстоутворювального чинника [169, ст. 23–30]. Синтаксичний міжфразовий зв'язок (МФЗ) – зв'язок між реченнями, коли одне з них є насиченням іншого в плані вираження (експліцитні МФЗ) або в плані змісту (імпліцитні МФЗ) [63]. Кількість міжфразових зв'язків може слугувати критерієм відбору речень у реферат, оскільки вимірює функціональну вагу речення в оригінальному тексті, яка є показником його значущості [155, ст. 27–31].

Текст як комунікативний знак володіє відносною комунікативною завершеністю і визначається конкретною темою, прагматикою і творчим задумом автора. Його структурні властивості настільки варіюються від одного тексту до іншого, що практично немає можливості говорити про наявність у мові формально-структурного інваріанту тексту. Відсутність у мові досить стійких формально-структурних характеристик тексту свідчить про те, що текст є собою мовленнєвою, а не мовною одиницею (у цьому плані показово, що при множинності типології текстів немає власне лінгвістичної типології) [44, ст. 18].

Однак із цього не випливає, що текст не може бути об'єктом лінгвістичного аналізу. Хоча текст і не може розглядатися як одиниця опису змісту, він складається з речень, здатних бути «мінімальним текстом». Це спонукає багатьох лінгвістів розглядати саме речення як одиницю, що має, як і текст, комунікативне призначення і виступає не тільки як мовленнєва, але й мовна одиниця. Речення виступає як одиниця тексту, що безпосередньо співвідноситься з мовним актом і разом з тим є мовним засобом вираження думок, ідей. Речення є значимою структурною текстовою та комунікативною одиницею [3].

Такі особливості речення, на думку багатьох сучасних лінгвістів, роблять його центральною структурною одиницею мови, створенню якої слугують, у кінцевому результаті, всі інші компоненти мовної системи.

На той факт, що речення виступає не тільки як мовленнєва, але й мовна одиниця, вказує той факт, що його структура, на відміну від структури тексту, завжди відповідає одній із формально-змістових моделей синтаксичної системи мови. Такі моделі описують семантичну структуру речення і типовий інформативний зміст, загальний для цілого класу однорідних речень. У кожного речення є початкова структура, яка може виглядати як підметно-присудкова структура або як конструкція підмет-присудок-додаток, або може мати предикатне ядро з валентностями, що заповнюються відповідними назвами. В той самий час у одному й тому ж реченні поєднуються декілька різних змістовних і формальних структур: предикатно-аргументна структура, суб'єктно-предикатна структура й актуалізаційні структури, які прив'язують речення до предметної ситуації, яке описується, і до ситуації висловлювання (інформаційна, ідентифікаційна, тематична, інтенціональна та ін.).

Кожна з таких структур виступає як спосіб «пакування» інформації, що передається за допомогою речення. Об'єднання в реченні декількох різнопланових, але при цьому мовних характеристик, що мають відповідні формальні засоби вираження, дозволяє розглядати речення як основну одиницю опису змісту тексту.

Звідси, з нашої точки зору, моделювання процесу реферування раціонально починати з побудови моделі реферату як набору можливих синтаксичних конструкцій, що зустрічаються в реферативних текстах.

Відомо, що синтаксична структура мови дозволяє звести безліч конкретних елементарних ситуацій до відносно невеликої кількості їх елементарних синтаксично-семантичних схем. Такі схеми закладені в самій природі речення. Речення містить у собі синтаксеми, що є елементарними значущими одиницями в складі синтаксичних конструкцій. За звичай це

значуще слово або ж поєднання значущого слова зі службовим у їх синтаксичному використанні.

При розробленні систем АР однією з ключових проблем була й залишається проблема аналізу семантичної структури вихідного тексту з метою визначення фактографічної інформації в межах тематики конкретного тексту з наступним її узагальненням і синтезом тексту реферату.

Наш підхід орієнтований на створення інтелектуальної системи автоматичного реферування. Попри цілковите усвідомлення всієї складності завдання основний акцент зроблено на спрощенні процедури реферування на початковому етапі досліджень за рахунок свідомо допущеної теоретичної та емпіричної неповноти. Однак зроблено це лише заради того, щоб надалі, відштовхнувшись від розуміння вивчених закономірностей, розширювати поле дослідження.

У межах такого підходу робота зі створення системи автоматичного реферування передбачає дослідження найбільш загальних закономірностей реферування, виражених у кінцевому продукті – рефераті.

У нашій роботі моделювання процесу реферування було зведено до кількох самостійних, однак взаємозумовлених завдань, розв'язання яких здійснювалося поетапно. На першому етапі було здійснено:

- 1) вивчення особливостей синтаксичної та семантичної структур реферату;
- 2) моделювання компресії на всіх рівнях;
- 3) побудова моделі уніфікованого реферату;
- 4) розробка алгоритму наповнення моделі реферату відповідною семантикою.

Наявність моделі реферату дозволяє забезпечити синтаксичну й семантичну правильність синтезованого реферату. Тоді як змістова повнота його залежить від рівня опрацювання питань змістового аналізу вихідного тексту й розробки механізму вилучення з нього необхідної для реферату інформації. Тому на другому етапі головне завдання звелось до:

1) розробки технології семантичного аналізу тексту для забезпечення глибокого проникнення в зміст тексту з метою трансформації його смислу в процесі реферування;

2) побудови семантико-контекстної моделі реферування, до якої входять модель заголовка й текстова база;

3) вибору засобів подання знань у системі автоматичного реферування.

Для розв'язання цього завдання необхідно було вивчити семантичні відношення вихідного тексту й реферату та побудувати модель подання знань у системі АР, що й стало об'єктом нашого дослідження.

Висновки

1. Сучасний етап у розвитку систем опрацювання інформації характеризується як етап формалізації і застосування знань.

2. Аналіз стану розробок у галузі автоматичного реферування дозволяє зробити висновок, що проблема удосконалення систем АР пов'язана з проблемою моделювання розуміння текстової інформації.

3. Розробники засобів автореферування все більше схиляються до гібридних систем, успішно поєднуючи статистичні методи й методи штучного інтелекту з застосуванням баз знань і онтологій.

4. Для автоматичного аналізу текстових даних усе частіше застосовуються методи Text Mining, які поєднують у собі технологічні й методологічні підходи аналізу змісту, комп'ютерної лінгвістики та штучного інтелекту у вигляді потужних сучасних систем, які включають елементи автореферування.

5. Задати єдиний алгоритм моделювання процесу розуміння текстової інформації неможливо. Кожен з розроблених алгоритмів висвітлює лише певні аспекти знань і в прикладному плані орієнтований на визначений клас завдань.

6. Основна проблема при моделюванні процесу реферування полягає в описуванні процедури стиснення інформації при переході від змістової структури тексту до змістової структури реферату.

II. ПОБУДОВА СИСТЕМИ АВТОМАТИЧНОГО РЕФЕРУВАННЯ З ОПЕРТЯМ НА ЗНАННЯ

2. 1. Моделювання компресії тексту в процесі реферування

Усі спроби автоматизації реферування, запроваджені впродовж останнього півстоліття, звелися однозначно до того, що використання виключно статистичних методів не побудує реальну систему реферування, тобто систему, яка надає винятково короткий виклад змісту, а не формує набір уривків тексту. Потрібні відповідні засоби подання знань – семантичні словники й онтології, що є базами знань певного виду. Тому на сьогодні в усьому світі ведеться широкомасштабна робота зі створення інтелектуальних систем реферування з опертям на знання [189, ст. 15–20].

Запропонований нами підхід є однією зі спроб створення такого плану системи АР. У межах пропонованого підходу понад десять років ведуться дослідження, що мають на меті моделювання процесу реферування, його формалізацію та створення на основі розроблених формалізмів системи автоматичного реферування (АР) з опертям на знання.

Роботу зі створення системи автоматичного реферування було розпочато з вивчення найбільш загальних закономірностей для всіх видів стиснення тексту, тобто свідомо припускалася теоретична й емпірична неповнота на початковому етапі, однак, лише з тією метою, щоб надалі, відштовхнувшись від розуміння вивчених закономірностей, розширювати сферу дослідження.

Моделювання процесу реферування, таким чином, було зведене до кількох самостійних, але взаємозумовлених завдань, розв'язання яких здійснювалося поетапно.

На першому етапі було проведено дослідження змістового наповнення рефератів, виділено змістові аспекти, що найбільш часто в них зустрічаються, й побудовано уніфікований реферат у вигляді набору речень із семантикою об'єкта, результату, мети й засобу.

Згодом було проведено дослідження й моделювання компресії (і узагальнення) у процесі реферування тексту на синтаксичному й семантичному рівнях.

Для розв'язання цих завдань необхідно було:

- розробити методичку формування ознак, що характеризують реферат як результат компресії (і узагальнення) інформації на синтаксичному й семантичному рівнях;

- провести дослідження синтаксичної структури реферативних речень і речень первинних текстів, і побудувати модель процесу стиснення на синтаксичному рівні;

- проаналізувати механізми стиснення на семантико-синтаксичному, власне семантичному й лексико-семантичному рівнях, побудувати модель процесу стиснення на семантичному рівні;

- побудувати модель реферату у вигляді типових для індикативних рефератів семантико-синтаксичних структур;

- виявити й класифікувати лексику, що використовується в індикативних рефератах;

- сформулювати правила наповнення синтаксичних конструкцій реферату поняттями, що узагальнюють семантику вихідного тексту.

Розроблена модель компресії лягла в основу першої версії комп'ютерної програми «АвтоРеферат». Аналіз результатів роботи даної програми продемонстрував правильність творення реферативних речень відповідно до розробленої моделі реферування, однак указав і на очікувану змістову неповноту цих речень і необхідність більш глибокого змістового аналізу первинного тексту, що лягло в основу завдання другого етапу досліджень.

Для розв'язання цього завдання слід було перейти до вивчення семантичних відношень вихідного тексту й реферату та побудови моделі демонстрації знань у системі АР у вигляді:

- 1) моделі заголовка, що являє собою концентроване вираження змісту вихідного тексту;

2) текстової бази, що являє собою «інформаційне ядро» тексту, яке містить інформацію, залежну від ситуації;

3) онтологій, що містять не залежну від ситуації інформацію: онтологій верхнього рівня, онтологій загальнонаукової лексики й онтологій предметних галузей.

Семантичні відношення вихідного тексту й реферату стали об'єктом дослідження на даному етапі.

Для опису проведеного дослідження в області семантичних відношень «Текст-Реферат» необхідно коротко описати результати, одержані на першому етапі аналізу, оскільки вони являють собою відправну точку всієї роботи.

Моделювання будь-якої системи починається з аналізу вимог, яким повинна відповідати система, що розробляється, з погляду її призначення й умов майбутнього використання, виділення основних об'єктів, які складають цю систему, та відношення між ними [43].



Рис. 2. 1 Об'єктна модель системи автоматичного реферування

Відправним пунктом у пізнанні змісту тексту є відповідь на запитання: що саме слід вважати за основний засіб вираження змісту документа – слово, окреме речення чи увесь текст загалом. Характер відповіді значною мірою буде зумовлювати вибір шляху та засобів вирішення поставленого завдання.

Відомо, що слову здебільшого не притаманна комунікативна завершеність. Його слід розглядати лише як складову, хоча й необхідну, частину тексту.

У підрозділі 1. 4 зазначалося, що основним засобом вираження змісту текстової інформації є речення. Текст не може розглядатися як одиниця опису змісту, оскільки становить скоріше мовленнєву, ніж мовну одиницю, про що свідчить відсутність у мові формально-структурного інваріанта тексту. Речення, що виступає не тільки як мовна, але й як мовленнєва одиниця, є складовою комунікативного акту і відповідно – носієм змісту. Додамо, перевага речення над текстом у даному розумінні полягає ще й у тому, що речення завжди можна подати у вигляді формально-змістової моделі, яка описує його семантичну структуру.

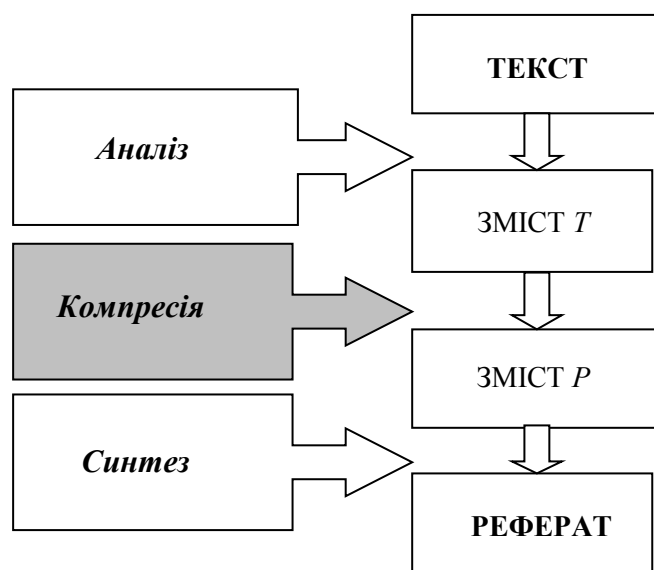


Рис. 2. 2 Етапи процесу реферування

Аналіз змісту вихідного тексту при реферуванні має свою специфіку, яка полягає в тому, що хоча тематика реферату й визначається тематикою вихідного тексту, проте реферування припускає не просто розпізнавання змісту тексту, а й подальшу його компресію. Причому остання здійснюється як на значеннєвому, так і на текстовому рівні. Із цього випливає, що аналіз змісту вихідного тексту слід проводити з урахуванням особливостей синтактико-семантичної організації реферату, що це, власне, і є кінцевою метою

процедури реферування. Звідси вихідним об'єктом дослідження стали синтаксичні й значеннєві особливості реферативних конструкцій.

На противагу елементарним одиницям тексту, слову чи реченню, реферат становить складний об'єкт, а саме – текст, що складається з набору речень певного типу.

Для формального опису такого об'єкта необхідно було провести аналіз синтаксичних конструкцій (СК), які зустрічаються в рефератах, і виділити множину типів СК реферативних речень для об'єднання їх у синтаксичну модель (шаблон) реферату (Додаток А).

Для формалізації виділеного об'єкта в роботі використовується метод, при якому об'єкт дослідження описується не всією притаманною йому інформацією, а набором ознак P , що характеризують (у нашому випадку) реферат R як результат компресії вихідного тексту T . Зв'язок цих ознак із рефератом визначається відношенням компресії – $K(P, R)$.

Поле визначення змінної R утворює множину конструкцій індикативного реферату певного типу:

[сирконстант] [предикат] [актант 1] [актант 2] ... [актант n]

Замінімо кожен елемент реферативної конструкції символом класу, до якого він належить. Здобутий ланцюжок описують синтаксичні конструкції речень індикативних рефератів:

$$PK: Sr - V - A_1 - A_2 - \dots - A_n,$$

де V – предикат, $A_1 A_2 \dots A_n$ – актанти, які є обов'язковими елементами РК, Sr – сирконстант, що є необов'язковим елементом РК.

Актанти $A_1 A_2 \dots A_n$ – іменні групи, що є неподільними семантичними одиницями й складаються з одного чи двох елементів:

$$N; N-Adj; Adj-N; N-N,$$

де N – іменник – може бути головним елементом актанта, якщо $N-Adj$ або $Adj-N$, або, якщо стоїть на першому місці в актанті, у випадку $N-N$ може бути залежним елементом, який стоїть на другому місці.

Adj – прикметник – може бути тільки залежним елементом актанта.

Поле визначення змінної P – множина ознак, що характеризують предикат, актанти й сирконстант. Вибір структури змінної P визначається зв'язками реферату з вихідним текстом.

Для реферату релевантними є синтаксичні, семантичні й частково граматичні зв'язки з вихідним текстом. Відповідно до цих типів зв'язків, структура змінної P має вигляд кортежу $\langle SS, S, G \rangle$, де SS – набір синтаксичних ознак $(ss_1, ss_2, \dots, ss_n)$, S – набір семантичних ознак (s_1, s_2, \dots, s_m) , G – набір граматичних ознак (g_1, g_2, \dots, g_k) .

Опис відношення компресії K зводиться, таким чином, до з'ясування відношень $K_1(R, SS)$ і $K_2(R, S)$, які характеризують усі типи впливу вихідного тексту на реферат (компресія на синтаксичному й семантичному рівнях), а також відношення $K_3(R, G)$, яке описує граматичні характеристики предикатів, актантів і сирконстантів, що входять у реферативне речення:

$$K = K_1(R, SS) K_2(R, S) K_3(R, G).$$

Модель реферату можна подати у вигляді набору реферативних конструкцій, які відображають усі названі типи компресії:

$$R = \{PK_1, PK_2, \dots, PK_n\}$$

Процес компресії, на наш погляд, є найважчим етапом процесу реферування, оскільки припускає згортання смислу шляхом пошуку найбільш об'ємистих засобів і форм подання інформації. При цьому зміст реферату повинен залишатися семантично адекватним і еквівалентним первинному документу. Моделювання саме цього етапу реферування дозволяє максимально інтелектуалізувати системи АР, оскільки він безпосередньо пов'язаний зі стисненням змісту тексту.

Процедура моделювання процесу компресії змісту при реферуванні складається з таких етапів:

- морфемний і морфологічний аналіз елементів синтаксичних конструкцій у складі індикативних рефератів і укладання повного списку морфологічних значень цих елементів (відношення K_3);

- аналіз синтаксичної структури реферативних конструкцій, подання основи моделі реферату у вигляді набору формалізованих типових синтаксичних конструкцій та укладання списку синтаксичних зв'язків між елементами цих конструкцій (відношення K_1);
- семантичний аналіз речень, що складають текст реферату, їх класифікація відповідно до семантичного значення (відношення K_2);
- лексико-семантичний аналіз текстів рефератів, класифікація загальнонаукової і термінологічної лексики.

Процес компресії при реферуванні нерозривно пов'язаний з процесом розуміння текстової інформації, оскільки він становить перехід від змістової структури тексту першоджерела до змістової структури реферату. Модель розуміння текстової інформації містить у собі модель знань (способів їхнього формалізованого представлення), що міститься в тексті. Це означає, що модель компресії змісту повинна являти собою лінгвістичну модель, яка включає в себе модель представлення знань.

Отже, важливим видається послідовний опис компресії тексту на всіх рівнях.

2. 1. 1. Компресія на морфологічному рівні

Морфологічні (граматичні й словотворчі) характеристики слів, які є елементами реферативних конструкцій (предикатів, актантів і сирконстантів), дозволяють аналізувати план змісту цих елементів з урахуванням особливостей плану вираження.

Пропонований у межах цього дослідження морфологічний аналіз зводиться до вивчення граматичних значень актантів, а також до часткового словотворчого аналізу предикатів на предмет наявності в них формальних ознак компресії.

Під морфологічним аналізом передбачається аналіз граматичних значень $G = \{g_1, g_2, \dots, g_k\}$ дієслів, що виступають у ролі предикатів V у СК індикативних рефератів, а також значень усіх актантів A і сирконстантів Sr , що входять до їх

складу. Словотворчий аналіз має на меті з'ясування структури словоформ, розглянутих ізольовано, тобто без залучення відомостей про їх навколишній контекст, тобто виявлення в дієсловах V префіксальних морфів із окресленим значенням «результату» та «повноти дії» – q [139].

Для опису граматичних значень слів, що входять до складу СК реферативного тексту, було виведено граматичні ознаки $G = \{g_1, g_2, \dots, g_n\}$ зі значеннями $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$: g_1 – частина мови зі значеннями: g_1^1 – іменник, g_1^2 – прикметник, g_1^3 – дієслово, g_1^4 – дієприкметник, g_1^5 – дієприслівник, g_1^6 – прислівник; g_2 – вид зі значеннями: g_2^1 – доконаний, g_2^2 – недоконаний; g_3 – стан зі значеннями: g_3^1 – активний, g_3^2 – пасивний; g_4 – час зі значеннями: g_4^1 – теперішній, g_4^2 – минулий, g_4^3 – майбутній; g_5 – особа зі значеннями: g_5^1 – перша, g_5^2 – друга, g_5^3 – третя; g_6 – число зі значеннями: g_6^1 – одиниця, g_6^2 – множина; g_7 – відмінок зі значеннями: g_7^1 – називний, g_7^2 – родовий, g_7^3 – орудний, g_7^4 – місцевий, g_7^5 – знахідний, g_7^6 – давальний (відмінкова парадигма подається не в традиційному порядку, а з огляду на частоту використання в реферативних конструкціях); g_8 – форма зі значеннями g_8^1 – повна, g_8^2 – коротка.

За цими ознаками граматичні правила формування всіх елементів реферативних конструкцій можна описати за допомогою граматичних предикатів Gr .

Граматичний предикат $Gr_1 = g_1^4 g_2^1 g_3^2 g_4^2 g_8^2$ описує правило формування граматичного значення V_1 , яке в нашому випадку здобуває ад'єктивне значення: той, що піддається дії в минулому й утримує результат цієї дії: *изучен, рассмотрен, сформулирован, продемонстрирован, проанализирован* [162, ст. 666].

Граматичний предикат $Gr_2 = g_1^3 g_2^2 g_3^2 g_4^1$ описує правило формування граматичного значення V_2 , яке має значення дії, реалізованої одночасно з моментом мови і такої, що прагне до досягнення результату: *изучается, рассматривается, формулируется, демонстрируется, анализируется*.

Головний елемент актанта може мати декілька граматичних значень.

Граматичний предикат $Gr_3 = g_1^1 g_7^1$ описує правило формування граматичного значення A_1 : *актуальность, функционирование, проблема*.

Оскільки значення відмінків формуються на основі їх синтаксичних функцій у реченні, тут суб'єктне значення (суб'єкт стану) називного відмінка (g_7^1) ускладнюється значенням об'єктним (об'єкт дії), тому що актант, який описується граматичним предикатом Gr_3 , входить до складу реферативної конструкції пасиву (підрозділ 2.2) [162, ст. 480–481].

Як головний, так і залежний елемент актанта може мати граматичні значення g_1^1, g_7^2 : *исследований, лечения, трансплантации*. Якщо іменна група актанта складається з двох іменників $N - N$, то залежний елемент завжди має такі самі граматичні значення, що описується предикатом $Gr_4 = g_1^1 g_7^2$. Головний елемент актанта також може мати значення $g_1^1 g_7^3$, наприклад: *методами, моделями, с помощью*, що описується предикатом $Gr_5 = g_1^1 g_7^3$.

У ролі головного й єдиного елемента актанта може виступати іменник зі значенням g_1^1 та g_7^4 , наприклад: *в мире, в Украине, в данной области*. У даному випадку місцевий відмінок із прийменником має диференційовано виражене просторове значення. Це правило описується предикатом $Gr_6 = g_1^1 g_7^4$.

У ролі залежного елемента актанта може виступати прикметник у повній формі з граматичними значеннями g_1^2 і g_8^1 , що описується предикатом $Gr_7 = g_1^2 g_8^1$.

Граматичний предикат $Gr_8 = g_1^6$ описує правило формування граматичного значення сирконстантів Sr у реферативних конструкціях, які заповнюються винятково прислівниками та позначають непроцесуальну ознаку дії (спосіб дії): *детально, подробно, поэтапно*, – і таким чином виконують функцію оцінної компресії у рефераті.

У результаті словотворчого аналізу дієслів і коротких прикметників – предикатів РК – ми виявили, що найчастіше в їхньому словотворі беруть участь префікси до-, из-, о-, при-, про-, раз-, с- [206]. Звичайно, існує множинність відтінків значень у кожного префікса, але тут спостерігається

спільна основа. Незважаючи на те, що в дієслівного префікса є основне значення, те, що робить його близьким із прийменником, дані префікси в цьому контексті мають [85]:

або семантичне значення q_1 = «результат дії»:

до- (доведен – доводиться), *при-* (приведен – приводиться), *про-* (проведен – проводиться, проаналізований), *с-* (сформульований, сведен – сводиться [162, ст. 361];

або q_2 = «повнота дії»:

из-(*ис-*) (измерен – измеряется), *о-*(*об-*) (описан – описывается), *раз-* (*рас-*) (рассмотрен – рассматривается) [101].

При цьому префікс зі значенням q_1 чи q_2 підсилює семантику результативності в V_1 (проаналізований, рассмотрен, обоснован) і додає V_2 відтінок значення – прагнення до досягнення результату (изучается, описывается, приводится).

2. 1. 2. Компресія на синтаксичному рівні

Для наукових текстів характерне використання невеликої кількості синтаксичних структур простих речень. Це пояснюється стилістичними особливостями даного виду текстів [208].

Прості речення, що входять до складу наукового тексту, незалежно від предметної галузі, представляють п'ять основних синтаксичних конструкцій, згідно з запропонованою Н. Ю. Шведовою градацією [162]:

$СК^1 : Part_{1коротк.} - N_1$

$СК^2 : Vf_{3s(pl)} - N_1$

$СК^3 : N_1 - N_2 - Adj_{1коротк.}$

$СК^4 : Vf_{1pl} - N_4$

$СК^5 : N_1 - N_2 - Vf_3$

Відповідно в рефератах використовуються лише дві синтаксичні структури речення [205]:

1. Пасивно-зворотна синтаксична конструкція CK_1 , в якій присудок виражений коротким пасивним дієприкметником:

$$CK_1: V_1 - A_1.$$

Наприклад: *Предложена методика; Получены зависимости; Введены определения; Рассмотрены вопросы; Установлены группы; Внимание уделено; Обоснована важность; Разработан метод.*

2. Синтаксична конструкція CK_2 , в якій присудок виражений дієсловом у 3-ій особі множини чи однини з постфіксом *-ся*:

$$CK_2: V_2 - A_1.$$

Наприклад: *Рассматривается необходимость; Исследуется структура; Сравниваются способы; Обсуждается вопрос; Освещаются результаты.*

Доцільність використання в рефератах саме даних двох синтаксичних структур спричиняється їх семантичним значенням. Як перша, так і друга конструкції мають таку семантичну схему: «відношення між суб'єктом його предикативною ознакою – станом як результатом дії».

Речення на зразок *Автора поддерживают коллеги* (автор отримує підтримку колег і колеги надають підтримку автору) і *Работу провели ученые* (робота була проведена вченими і вчені провели роботу), що мають тотожне семантичне значення, не зустрічаються в рефератах, оскільки в них присутній реальний суб'єкт, який виступає в ролі синтаксичного об'єкта [162, ст. 256]. А це, як ми з'ясували в процесі дослідження, в рефератах опускається задля стислості (підрозділ 2.2).

Прості речення в індикативних рефератах найчастіше поширені постпозитивним дієприкметниковим зворотом. Дієприкметник при підметі-іменнику, що входить до складу відокремленого звороту, звичайно позначає ознаку, постійно властиву суб'єктові, наприклад: *предлагается подход, основанный на стандартизации структуры; рассмотрены уравнения, описывающие процессы.* У цьому випадку семантика відокремленого дієприкметникового звороту еквівалентна семантиці простих речень у рефератах і виражає ознаку, як результат дії. Це дозволило відійти від

використання дієприкметникових зворотів у синтаксичних конструкціях типової моделі реферату з метою спрощення процесу автоматичного реферування і замінити їх еквівалентними простими реченнями, наприклад: *предлагается стандартизация структуры, рассмотрены описания процессов.*

Таким чином, компресія на синтаксичному рівні в процесі реферування відбувається шляхом використання синтаксичних конструкцій простих речень із семантичним значенням результативності (рис. 2. 3).

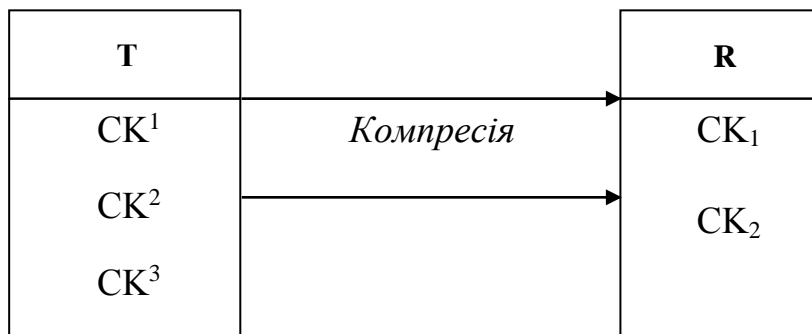


Рис. 2.3 Компресія на синтаксичному рівні

У кількісному відношенні індикативний реферат утворює набір від одного до трьох (найчастіше з двох) речень однієї з синтаксичних конструкцій CK_1 чи CK_2 , що зустрічаються в рефератах.

Реферат у вигляді $R_1=\{CK_1CK_1\}$: «Дана токсиколого-гигиеническая и медико-биологическая характеристика поверхностно-активных веществ. Обобщены данные литературы и результаты собственных исследований медико-биологических аспектов проблемы загрязнения окружающей среды поверхностно-активными веществами» чи $R_2=\{CK_2CK_2\}$: «Анализируется инвестиционная привлекательность различных регионов Украины в контексте современных экономических преобразований. Формулируется система практических рекомендаций для более успешного привлечения потенциальных инвесторов» є більш вдалим, аніж припустімо: $R=\{CK_2CK_1\}$: «Рассматривается системный подход к обоснованию и моделированию социетальной психики неслучайных групп. Рассмотрен интегральный тип информационного метаболизма нации» чи $R=\{CK_1CK_2\}$: «Обоснована

важність изучения смисла на словообразовательном уровне. Рассматривается необходимость математического описания смисла».

Формальне представлення реферативних конструкцій у вигляді

$CK_1: V(Gr_1) - A_1(Gr_3)$ і $CK_2: V(Gr_2) - A_1(Gr_3)$ є тільки основою парадигм реферативних речень, тому що речення, що входять до його складу, містять потенційно до семи іменних актантів.

Існують варіанти поширення синтаксичної конструкції CK_1 , що найчастіше застосовуються (Додаток Б) до типового з трьохслівним чи більш довгим ланцюжком словоформ іменників (жирним шрифтом виділені обов'язкові елементи):

$CK_1: Sr(Gr_8) - V(\mathbf{Gr}_1) - A_1((Gr_7)(\mathbf{Gr}_3)) - A_2((Gr_7)(Gr_4)) -$
 $A_3((Gr_7)(Gr_4)) - A_4((Gr_7)(Gr_5)) - A_5((Gr_7)(Gr_4)) -$
 $A_6((Gr_7)(Gr_4)) - A_7((Gr_7)(Gr_4)).$

Висока частота вживання в РК іменників у родовому відмінку (Gr_4) пояснюється тим, що родовий відмінок легко вступає в підрядні зв'язки з іменниками, виконуючи роль атрибута. В наукових текстах родовий відмінок дозволяє уникнути неоднозначності, неточності, робить мову більш чіткою, заощаджує місце й час [73].

Синтаксична конструкція CK_2 зустрічається в реферативних текстах трохи частіше, ніж конструкція CK_1 , однак має менше варіантів поширення (Додаток В). Тому вона була уніфікована за рахунок поширення синтаксичної структури CK_2 до типової:

$CK_2: Sr(Gr_8) - V(\mathbf{Gr}_2) - A_1((Gr_7)(\mathbf{Gr}_3)) - A_2((Gr_7)(Gr_4)) - A_3((Gr_7)(Gr_4)) -$
 $A_4((Gr_7)(Gr_4)) - A_5((Gr_7)(Gr_4)) - A_6((Gr_7)(Gr_4)) - A_7((Gr_7)(Gr_4))$ [109].

Між елементами двоскладних актантів (іменної групи) можливі два типи синтаксичного зв'язку: ss_1 – узгодження і ss_2 – керування. Зв'язок ss_1 виражає означальні відношення між іменником у ролі головного елемента іменної групи й прикметником, що виконує функцію визначення в ролі залежного

елемента (*словообразовательное значение, современными методами, финансовых ресурсов*):

$$ss_1 : \{Adj - N\}$$

і описується синтаксичним предикатом Sn_1 :

$$Sn_1 = Adj(Gr_7)N(Gr_3 \vee Gr_5 \vee Gr_6)$$

Зв'язок ss_2 (керування) установлюється як між елементами ланцюжка іменників у складі іменної групи актанта, так і між іменниками, що є головними елементами різних актантів. Тут мова йде про сильне керування, при якому виникають заповнюючі відношення. При цьому головний елемент зв'язку (іменник) може мати форму називного, родового, орудного чи місцевого відмінка, а залежний (іменник) – форму родового чи місцевого відмінка (*база данных, язык запросов, процедуры вывода*):

$$ss_2 : \{N - N\}$$

і описується синтаксичним предикатом Sn_2 :

$$Sn_2 = N(\underset{i=3}{\overset{4}{V Gr_i}}) N(Gr_4 \vee Gr_6), \text{ де } \underset{i=3}{\overset{4}{V Gr_i}} = Gr_3 \vee Gr_4$$

Синтаксичні зв'язки всередині актантів можна умовно назвати зв'язками нижнього рівня, тоді як до верхнього рівня відносяться синтаксичні зв'язки між актантами в складі реферативної СК. Тут, крім описаних вище зв'язків ss_1 і ss_2 , спостерігаються зв'язки: ss_3 – керування, ss_4 – прилягання і ss_5 – узгодження. Причому синтаксичним зв'язком між двоскладними актантами вважаються зв'язки між їх головними елементами.

Зв'язок ss_3 установлюється між присудком (V) і підметом (Sbj), які утворюють предикативний центр речення (*рассматривается необходимость, анализируется модель, описана реализация, обоснована актуальность*):

$$ss_3 : \{V - Sbj\}$$

і описується синтаксичним предикатом Sn_3 :

$$Sn_3 = \underset{i=1}{\overset{2}{V(V Gr_i)}} N(Gr_3).$$

Зв'язок ss_4 (прилягання) в реферативних СК установлює:

1) означальні відношення між дієсловом-присудком (V) і прислівником-обставиною (Sr) (*особо отмечается, подробно описаны, детально анализируется*): $ss_4: \{Sr - V\}$,

2) означальні відносини між дієсловом-присудком (V) та іменником (N^o) в орудному відмінку зі значенням способу (*описан моделями, описан с помощью, анализируется методами*): $ss_4: \{V - N^o\}$,

3) означальні відносини між дієсловом-присудком (V) та іменником (N^p) у родовому відмінку зі значенням мети (*для поиска и обработки охарактеризованы, для распределения сформулированы, для привлечения формулируется*): $ss_4: \{V - N^p\}$ і описується синтаксичним предикатом Sn_4 :

$$Sn_4 = Sr(\underset{i=1}{\overset{2}{Gr_8}})V(\underset{i=1}{\overset{2}{VGr_i}}) \vee V(\underset{i=1}{\overset{2}{VGr_i}})N(\underset{i=1}{\overset{2}{Gr_5}}) \vee V(\underset{i=1}{\overset{2}{VGr_i}})N(\underset{i=1}{\overset{2}{Gr_4}}).$$

Реферативні СК найчастіше поширені однорідними членами речення чи, як їх називають, сурядним рядом – рядом словоформ, об'єднаних сурядним зв'язком. Сурядний ряд у реферативних СК може займати позицію присудка, підмета, додатка або означення, наприклад:

Формулируется и доказывается принцип суперпозиции кратчайших предикатов.

Показаны актуальность и значимость дальнейшей перспективы трансплантации печени.

Анализируется необходимость формирования системы ипотечного, товарного и сезонного кредитования под залог сельскохозяйственной продукции.

Члени сурядного ряду завжди розташовуються контактено, при цьому самі можуть бути поширені, тобто мати при собі залежні від них словоформи, що виступають у ролі означення чи додатка:

Представлены методы купирования ангинозного приступа и неотложные мероприятия при нестабильной стенокардии.

Словоформи, що складають сурядний ряд, граматично пов'язані з яким-небудь одним, спільним для них членом. Це може бути присудок, підмет чи додаток:

Подробно описаны особенности отбора доноров и реципиентов, этапы хирургического вмешательства, анестезиологическое пособие, возможные осложнения.

Рассматривается проблема обоснования и моделирования социальной психики несчастных групп.

Сурядні ряди, що поширюють прості речення РК, є відкритими, тобто складаються з потенційно необмеженого числа членів, а отже, порядок послідовності членів даного ряду – вільний. Показником сурядного зв'язку виступає простий сполучник «и» чи інтонація (кома в письмових текстах). Сурядні відносини між словоформами всередині ряду є відношеннями поєднання (ss_5) синтаксично недиференційованих членів (члени ряду цілком рівноправні): $ss_5 : \{V - V\}$; $ss_5 : \{N - N\}$; $ss_5 : \{Adj - Adj\}$; $ss_5 : \{Sr - Sr\}$ і описується синтаксичним предикатом Sn_5 :

$$Sn_5 = V(VG_i) V(VG_i) v N(VG_i) N(VG_i) v Adj(G_7) Adj(G_7) v Sr(G_8) Sr(G_8).$$

$\begin{matrix} 2 & 2 & 6 & 6 \\ i=1 & i=1 & i=3 & i=3 \end{matrix}$

Зв'язок ss_5 є зв'язком нижнього рівня, якщо він поєднує однорідні залежні елементи всередині актанта, оскільки прикметник не може бути окремим актантом, наприклад: *ипотечное, сезонное и товарное кредитование, паллиативное и симптоматическое лечение*. В усіх інших випадках зв'язок ss_5 є зв'язком верхнього рівня, що поєднує актанти в складі СК.

Наявність сурядних рядів ускладнює синтаксичну структуру речення, однак робить реферативну конструкцію в цілому більш інформативною.

Таким чином, правила побудови синтаксичних зв'язків у реферативних реченнях описуються предикатом Sn :

$$Sn = V(Sn_i).$$

$i=1$

Отже, проведене дослідження показало, що компресія у процесі реферування на синтаксичному рівні полягає в побудові двох типових реферативних синтаксичних конструкцій CK_1 і CK_2 , які володіють семантикою результативності, характеризуються п'ятьма типами синтаксичних зв'язків усередині й між актантами та предикатами, що можна подати у вигляді синтаксичного відношення $K_1: (CK_1 \vee CK_2) Sn$.

2. 2. Моделювання компресії змісту тексту

Семантичне відношення $K_2(R,S)$ компресії змісту в процесі реферування описується відповідними предикатами: предикат SSn описує правила побудови семантичної структури речень, що входять до складу РК; предикат Sm описує правила формування семантичних значень речень у складі РК; предикат Ls описує правила класифікації лексем, що беруть участь у заповненні актантних структур реферативних конструкцій.

Кожна одиниця в синтаксисі – як сама конструкція, так і одиниця, що бере участь у її побудові, – має своє мовне значення (мовні значення). Для синтаксису істотним є як лексичне, так і граматичне значення слова. Цими значеннями регулюється участь слова (в тій чи іншій його формі або формах) у побудові речення. Тут складно взаємодіють морфологічне значення форми слова та його синтаксичних зв'язків. Найбільш очевидна така взаємодія у сфері відмінкових значень і у сфері значень дієслівних форм. Значення словосполучення формується як відношення його стрижневого й залежного компонентів; у реченні це значення може ускладнюватися й видозмінюватися під впливом конкретного інформативного завдання.

Особливо складна організація змістовної сторони речення. Просте речення містить у собі кілька значень різного ступеня абстракції. Граматичним значенням кожного речення є *предикативність*, тобто виражене спеціально для цього призначеними формальними засобами відношення того, про що повідомляється, в той чи інший час. Кожному типу речень різної граматичної будови властива своя семантика, що формується граматичними значеннями компонентів абстрагованого зразка такого речення, синтаксичним відношенням цих компонентів один до одного й правилами їх лексичного наповнення. В реченні, яке побудоване за структурною схемою і зберігає в собі семантику цього зразка, міститься також визначений зміст, що утворюється шляхом взаємодії семантики цього абстрагованого зразка з лексичними значеннями слів, що ввійшли до речення [162, ст. 11–12]. Одиниці значення, які називаються семантичними компонентами речення і формуються

на основі взаємовідношення синтаксичних значень його членів і лексичних значень слів, утворюють семантичну структуру речення. Центральними категоріями семантичної структури речення є: предикативна ознака (*Prd*), суб'єкт (*Sbj*) та об'єкт (*Obj*).

Хоча існує думка про те, що носієм предикативної ознаки може бути будь-яке слово в реченні залежно від логічного наголосу (логічний наголос – це виділення у вимові одного зі слів речення для посилення його змістового навантаження), ми дотримуємося класичного підходу щодо присудка як носія даної ознаки, оскільки маємо справу з письмовим текстом, а не усним мовленням [65].

У реферативних конструкціях суб'єктна ознака не виражена, оскільки синтаксична структура РК є пасивною без указівки на суб'єкт предикативної ознаки. Тому предикативна ознака в СК рефератів є безсуб'єктною.

Автор оригінального тексту (статті), суб'єкт дії, передбачається, але не називається в текстах індикативних рефератів для спрощення СК за рахунок стислості. А от граматичний суб'єкт, насправді, є реальним об'єктом, що цілком пояснюється пасивною конструкцією [208]:

рассматривается, предлагается, исследуется, сравниваются, – что? – необходимость, новая наука, структура, различные способы.

Наявність *Prd* і *Obj* у CK_1 і CK_2 індикативних рефератів обов'язкова, при цьому граматичний суб'єкт (підмет) є першим актантом, а кількість додаткових іменних актантів не перевищує шести. Наявність *Sr* у CK_1 і CK_2 індикативних рефератів свідчить про наявність суб'єктивної авторської оцінки, оскільки сирконстант *Sr* завжди виражений прислівником з оціночним значенням. Таким чином, семантична структура реферативних СК може бути представлена у вигляді:

$$K_s : Sr - Prd - Obj - A_2 - A_3 - A_4 - A_5 - A_6 - A_7 ,$$

де $Prd = V_i$, $Obj = A_1$.

Таким чином, стиснення інформації на семантико-синтаксичному рівні в процесі реферування відбувається за рахунок відсутності в СК реферативного тексту реального суб'єкта дії, що описується семантико-синтаксичним предикатом SSm

$$SSm = Sr Prd Obj \bigvee_{i=2}^7 A_i,$$

Компресія на семантичному рівні відбувається шляхом вживання в РК речень з визначеним семантичним значенням. Речення, що входять до складу реферативного тексту, незалежно від синтаксичної структури можуть мати два основних метазначення s_1 – об'єкт і s_2 – результат:

$$CK^o = CK(s_1); CK^p = CK(s_2).$$

CK^o описує об'єкт (чи предмет) дослідження оригінального тексту, CK^p представляє результати дослідження, описані в першоджерелі. Семантичне значення CK у складі індикативного реферату залежить від кількості речень у рефераті. Якщо реферат складається з двох речень, то:

$$R = \{CK_1^o, CK_1^p\} \text{ чи } R = \{CK_2^o, CK_2^p\},$$

наприклад:

Рассмотрены основные аспекты пересадки печени (s_1). Доказаны актуальность и значимость дальнейшей перспективы трансплантации печени (s_2).

Рассматриваются основные аспекты пересадки печени (s_1). Доказывается актуальность и значимость дальнейшей перспективы трансплантации печени (s_2).

У тих випадках, коли реферат представлений одним реченням, це речення описує результат дослідження, при цьому містить указівку на об'єкт:

$$R = \{CK_1^p\} \text{ чи } R = \{CK_2^p\},$$

наприклад:

Предлагается оригинальный подход (s_2) к построению описания процесса понимания естественно-языкового текста (s_1).

При цьому індикативна функція цілком виконується, оскільки в реченні є посилання на об'єкт і результат.

У тих випадках, коли реферативний текст містить три речення, два речення описують об'єкт, одне – результат, або навпаки:

$$R=\{CK_1^o, CK_1^o, CK_1^p\}, R=\{CK_1^o, CK_1^p, CK_1^p\}$$

$$R=\{CK_2^o, CK_2^o, CK_2^p\}, R=\{CK_2^o, CK_2^p, CK_2^p\},$$

наприклад:

Представлено современное определение нестабильной стенокардии (s_1). Сформулированы основные положения интегральной соционики (s_2). Показаны методы описания типов этносов и государств (s_1).

Порядок уживання речень CK^o і CK^p у РК не відіграє, на нашу думку, принципової ролі, хоча передування опису об'єкта CK^o представленню результату CK^p видається кращим з погляду логічного подання інформації.

Речення CK^p із семантикою результату можуть містити додаткові актантні ланцюжки з метазначенням s_3 – *мета*:

Анализируется инвестиционная привлекательность различных регионов Украины (s_1). Формулируется система рекомендаций (s_2) для привлечения потенциальных инвесторов (s_3).

Рассмотрены информационные аспекты стратегического управления предприятием (s_1). Для поиска и предварительной обработки стратегической информации (s_3) охарактеризованы свойства компьютерного агента (s_2).

Як речення CK^o , так і речення CK^p можуть містити додаткові актантні ланцюжки зі значенням s_4 – *засіб*:

Проанализированы соционические психотипы (s_1) по типологии Хейманса-Ле Сенна (s_4). Приведены результаты практической работы (s_2).

Методами этносоционики (s_4) проанализирована модель интегрального типа США (s_2). Рассмотрены психоинформационные взаимодействия США (s_1).

Формально актантні ланцюжки з метазначенням *мети* розпізнаються за наявністю прийменника «для» перед першим словом з граматичною ознакою

g^2 у ланцюжку актанта. Актантні ланцюжки з метазначенням *засобу* можна формально розпізнати за відмінковою ознакою g^3 першого слова актанта. У синтаксичній моделі реферативної СК, представленій у вигляді дерева залежностей, актантні ланцюжки з метазначенням *мети* й *засобу* утворюють окрему гілку.

Таким чином, власне семантичними ознаками S реферату R є:

$$S = (s_1, s_2, s_3, s_4),$$

де s_1 – об'єкт дослідження,
 s_2 – результат дослідження,
 s_3 – мета дослідження,
 s_4 – засіб дослідження.

А предикат, який описує правило формування семантичного значення речень у складі реферату має вигляд:

$$Sm = s_1s_2(s_3 \vee s_4).$$

Моделювання компресії в процесі реферування на лексико-семантичному рівні передбачає семантичну класифікацію дієслівних та іменникових актантів, що входять до складу речень РК наукових текстів.

У таких конструкціях дієслово є основною одиницею, яка визначає його значеннєві зв'язки з актантами, що заповнюються іменниками (чи іменниковими групами), і сирконстантами, що заповнюються прислівниками.

Заповнення сирконстантів є виключно інтелектуальним процесом (виконується лише людиною), оскільки прислівники, що їх заповнюють, є носіями семантики оцінювання.

Лексеми, що заповнюють предикат РК, класифікуються за метазначеннями m_1, m_2, m_3 на три класи: дієслова зі значенням m_1 – форма представлення інформації в тексті, m_2 – етап роботи, m_3 – порівняльне оцінювання.

До першого класу $V(m_1)$, де m_1 – форма представлення інформації в тексті, увійшли лексеми, які можна вважати контекстними синонімами у функції предиката.

$V(m_1) = \{вводит\ь, дават\ь, демонст\r{и}риват\ь, излагат\ь, описыват\ь, освещат\ь, отмечат\ь, показыват\ь, предлагат\ь, представлят\ь, приводит\ь, рассматриват\ь\}$.

До другого класу $V(m_2)$, де m_2 – етап роботи, потрапили лексеми, що володіють метатекстовим значенням – опис визначених етапів виконаної роботи:

$V(m_2) = \{анализироват\ь, внедрят\ь, выполнят\ь, доказыват\ь, измерят\ь, изучат\ь, использовать, исследоват\ь, находит\ь, получают, обобщат\ь, обосновыват\ь, обсуждат\ь, оцениват\ь, перечислят\ь, разрабатыват\ь, реализовыват\ь, решат\ь, указыват\ь, устанавливат\ь, формулироват\ь, характеризоват\ь\}$.

До третього класу $V(m_3)$ увійшли лексеми, що мають метатекстову семантику оцінки: m_3 – порівняльне оцінювання. Оцінювання проводиться укладачем реферату на основі порівняння результатів даної роботи з попередніми публікаціями:

$V(m_3) = \{добавлят\ь, доводит\ь, дополнят\ь, изменят\ь, исправлят\ь, модифицироват\ь, сравниват\ь, улучшат\ь\}$.

Використання в реферативному тексті лексем із третьої групи демонструє наявність у ньому суб'єктивної оцінки автора реферату. Абсолютно очевидно, що при створенні автоматичного реферату без участі людини неможливе внесення в текст реферату суб'єктивної оцінки.

Другою ознакою класифікації предикатів є граматична ознака. В ролі присудка реферативної СК може виступати дієслово зі значенням $Gr_2 = g_1^3 g_2^2 g_3^2 g_4^1$ (*рассматривается*) і скорочена форма дієприкметника з граматичним значенням $Gr_1 = g_1^4 g_2^1 g_3^2 g_4^2 g_8^2$ (*рассмотрен*) (підрозділ 2.1), які є семантично взаємозамінними в тексті індикативного реферату.

Таким чином, процес розбиття множинності предикатів РК на класи відбувався в два етапи (рис. 2.4):

1. Розподіл лексем, що виступають у ролі предикатів реферативних СК за метазначеннями m_1, m_2, m_3 на класи: $V(m_1), V(m_2)$ чи $V(m_3)$.

2. Розподіл лексем, що виступають у ролі предикатів реферативних СК, за граматичною ознакою G на класи:

$$V(Gr_1) = V(g_1^4 g_2^1 g_3^2 g_4^2 g_8^2) \text{ чи } (Gr_2) = V(g_1^3 g_2^2 g_3^2 g_4^1).$$

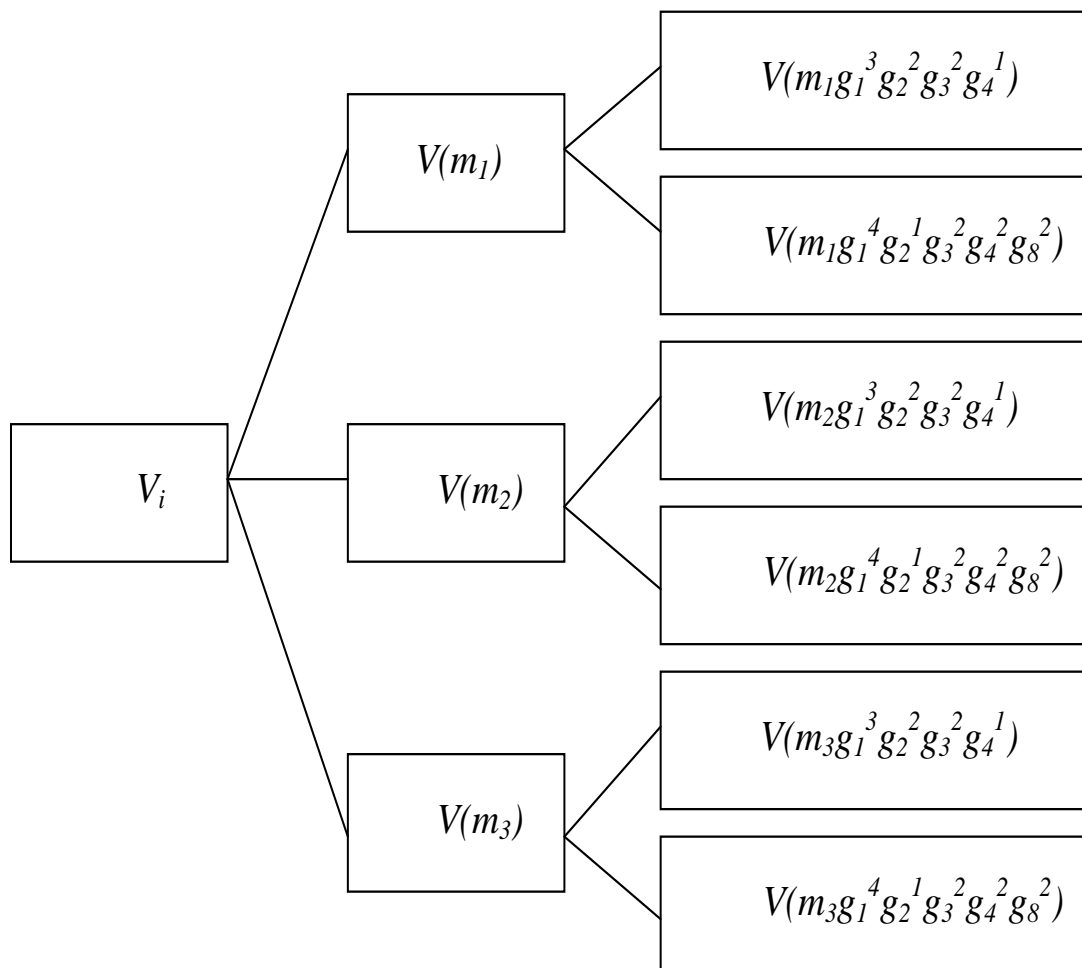


Рис. 2. 4 Лексико-семантична класифікація предикатів РК

У результаті класифікації було сформовано онтологію-словник предикатів реферативних конструкцій $O = \langle \{V\}, \{\}, \{\} \rangle$, що містить у собі скінченну множинність концептів-предикатів із загальнонаукової галузі:

$V(m_1 Gr_1) = \{ \text{анализируется, излагается, вводится, демонстрируется, описывается, освещается, отмечается, показывается, предлагается, приводится, рассматривается} \}$.

$V(m_1Gr_2) = \{проанализирован, введен, продемонстрирован, изложен, описан, освещен, отмечен, показан, предложен, представлен, приведен, рассмотрен\}$.

$V(m_2Gr_1) = \{внедряется, выполняется, доказывается, измеряется, изучается, используется, исследуется, обобщается, обосновывается, обсуждается, оценивается, перечисляется, разрабатывается, реализовывается, решается, формулируется, указывается, устанавливается, характеризуется, сводится\}$.

$V(m_2Gr_2) = \{внедрен, выполнен, доказан, измерен, изучен, использован, исследован, обобщен, обоснован, оценен, перечислен, разработан, реализован, решен, сформулирован, указан, установлен, охарактеризован, сведен\}$.

$V(m_3Gr_1) = \{добавляется, доводится, дополняется, изменяется, исправляется, модифицируется, сравнивается, улучшается\}$.

$V(m_3Gr_2) = \{добавлен, доведен, дополнен, изменен, исправлен, модифицирован, улучшен\}$.

Правило розподілу предикатів у реферативних реченнях на класи описуються предикатом L_{SV} :

$$L_{SV} = V(\bigvee_{i=1}^3 \bigvee_{j=1}^2 Gr_j).$$

У свою чергу іменникові актанти A заповнюються іменниковими групами, що складаються з іменників, за можливістю поширених залежними від них прикметниками (підрозділ 2.1). За класифікацію іменникових груп можна вважати класифікацію іменників N , що є головними елементами іменникових груп. Класифікація іменників N , що входять до складу іменникових груп, як і класифікація предикатів, на першому етапі відбувалася за семантичною ознакою. Лексичні значення іменників тут поєднуються метазначеннями $m_4 - m_9$, що дозволяє розбити всю множинність іменників N на шість класів іменників $N(m_4)$, $N(m_5)$, $N(m_6)$, $N(m_7)$, $N(m_8)$, $N(m_9)$ зі

значеннями m_4 – об'єкт, m_5 – процес, m_6 – ознака, m_7 – мета, m_8 – засіб, m_9 – місце.

Далі іменники були розподілені на три класи за ознакою приналежності лексеми до одного з трьох словників певної лексики:

ls_1 – термінологічна лексика (словник ПГ),

ls_2 – загальнонаукова лексика (словник ЗН),

ls_3 – загальноповсякденна лексика (словник ЗВ).

Те ж саме було зроблено зі словниковим складом індикативних рефератів даних текстів.

У процесі статистичного аналізу досліджуваних текстів до термінологічної лексики (ls_1) відносилися терміни, що зустрічаються в словнику даної ПГ [182; 110]. Термінологічне поняття представлене іменниковою групою, в якій іменник може мати залежний елемент.

До загальнонаукової лексики (ls_2) увійшли ті іменники (іменникові групи) з оригінальних текстів і текстів їх індикативних рефератів, які можуть зустрічатися в текстах будь-якої предметної галузі.

Таким чином була створена вироджена онтологія, яка являє собою словник загальнонаукової лексики російської мови, до якої увійшли лексеми трьох класів відповідно до метазначення (m_4, m_5, m_6) (Додаток В). Однак слід зазначити, що такі онтології-словники мають обмежене використання, оскільки не вводять експліцитного змісту термінів.

Лексеми, що не зустрічаються в словниках ПГ чи ЗН, вважалися приналежними до загальноповсякденної лексики (ls_3).

При подальшій класифікації іменникових груп РК з'ясувалося, лексеми яких словників є носіями метазначень $m_4 - m_9$ у реферативному тексті.

Носіями метазначення m_4 , що класифікується в онтології верхнього рівня як об'єкт, можуть бути загальнонаукові й термінологічні поняття, наприклад:

$N(m_4ls_1)$ – терміни з метазначенням «об'єкт»: *модель языка, семантическая модель, лексическое значение, система управления базами*

данных, язык запросов, информационно-поисковая система, база данных; тарифные ставки, себестоимость, экономические реформы, товарное производство, рыночная платежеспособность, экспортное производство, инвестиционный потенциал.

$N(m_4ls_2)$ – загальнонаукові поняття з метазначенням «об’єкт»: *алгоритм, аспект, вопрос, концепция, метод, модель, определение, положение, признак, принцип, причина, проблема, процедура, рекомендация, свойство, система, структура, тип.*

Носіями метазначення m_5 , що класифікується в онтології верхнього рівня як процес, можуть бути також загальнонаукові й термінологічні поняття, наприклад:

$N(m_5ls_1)$ – терміни з метазначенням «процес»: *медикаментозное лечение, купирование, трансплантация; инвестирование, кредитование;*

$N(m_5ls_2)$ – загальнонаукові поняття з метазначенням «процес»: *достижение, изучение, использование, исследование, обоснование, обработка, описание, организация, осуществление, построение, представление, преобразование, развитие, распределение, распространение, стандартизация, усовершенствование, формирование.*

Метазначення m_6 , m_7 , m_8 у реферативних текстах мають тільки загальнонаукові поняття, наприклад:

$N(m_6ls_2)$ – загальнонаукові поняття з метазначенням «ознака»: *актуальность, адекватность, важность, необъяснимость, популярность, правомерность, эффективность;*

$N(m_7ls_2)$ – загальнонаукові поняття з метазначенням «мета»: *для описания, для построения, для распределения, для улучшения;*

$N(m_8ls_2)$ – загальнонаукові поняття з метазначенням «засіб»: *методами, моделями, с помощью;*

Носієм метазначення m_9 у реферативних текстах є загальнонаукова лексика зі значенням місця проведення дослідження чи використання його результатів, наприклад:

$N(m_9ls_3)$ – загальноновживані поняття з метазначенням «місце»: *город, область, регион, Украина, Европа, мир.*

На третьому – заключному етапі класифікації іменникових груп, що беруть участь у заповненні актантів РК, відбувався поділ іменників за граматичною ознакою Gr залежно від тієї граматичної форми, в якій лексема з даними ознаками m і ls може вживатися в СК реферативного тексту.

Так поняття з метазначеннями: «об’єкт», «процес» та «ознака» можуть виступати у формі як називного, так і безприйменникового родового відмінка, тоді як поняття з метазначенням «мета» уживаються тільки у формі родового відмінка з прийменником, поняття з метазначенням «засіб» – у формі орудного, а поняття з метазначенням «місце» – місцевого відмінка.

Таким чином, класифікація іменникових груп відбувалась також у три етапи (рис. 2.5).

1. Поділ іменників, які входять до складу іменних груп реферативних СК за семантичною ознакою, на класи:

$N(m_4), N(m_5), N(m_6), N(m_7), N(m_8)$ чи $N(m_9)$.

2. Поділ за ознакою приналежності до лексики на класи:

$N(ls_1), N(ls_2), N(ls_3)$.

3. Поділ за граматичною ознакою на класи:

$N(Gr_3), N(Gr_4), N(Gr_5), N(Gr_6)$.

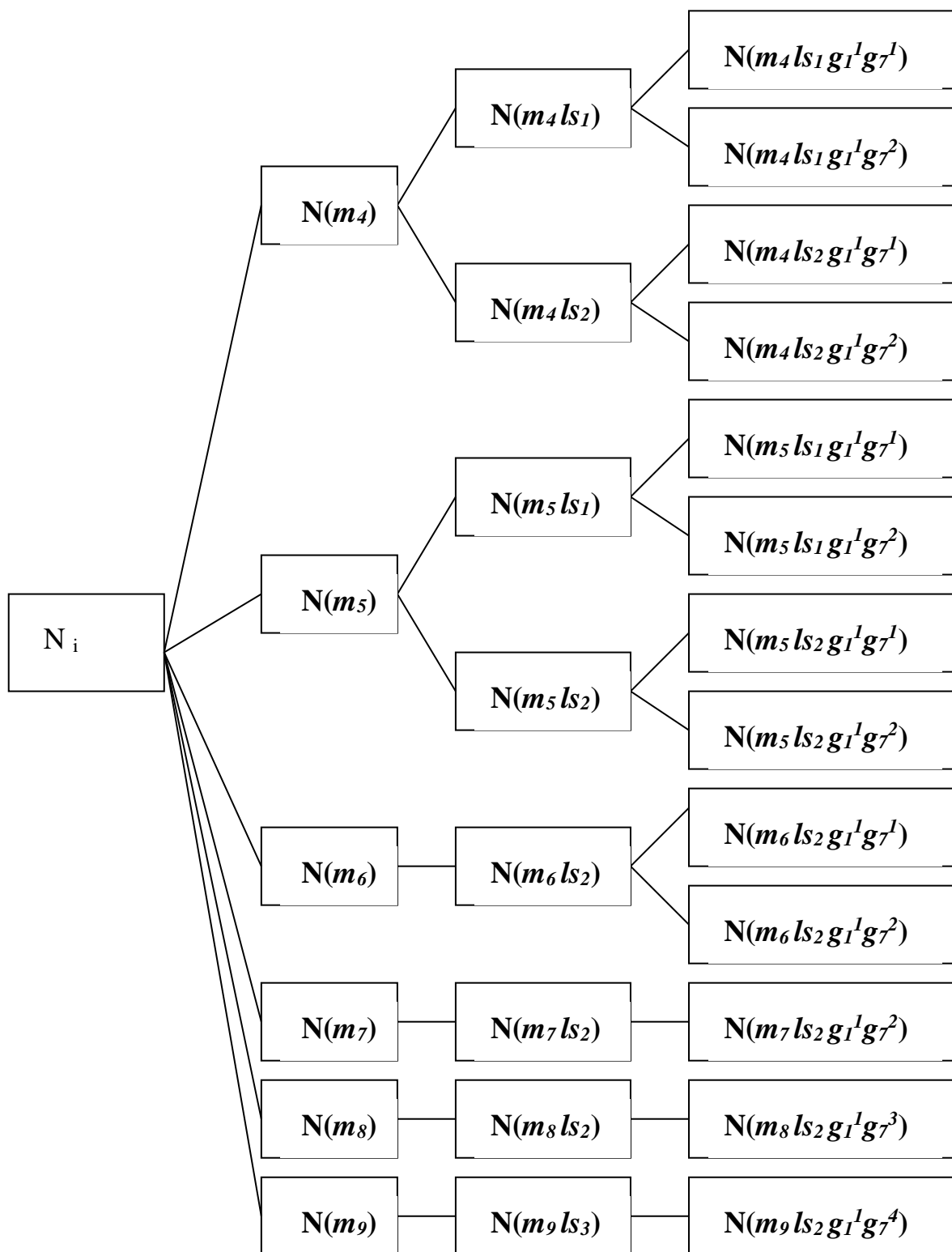


Рис. 2. 5 Лексико-семантична класифікація іменникових груп РК

Отже, правило розподілу лексем на класи згідно з набором метазначень лексем РК, що входять до онтології верхнього рівня (словника категорій), описуються предикатом Ls_N :

$$Ls_N = \bigvee_{i=4}^9 m_i (\bigvee_{j=1}^3 ls_j),$$

де m_i – метазначення $m_4 – m_9$; ls_j – словники термінологічної, загальнонаукової і загальноповживаної лексики. Семантичне відношення $K_2(R,S)$ компресії змісту має вигляд (підрозділ 2.1):

$$K_2 = SSm Sm (Ls_V \vee Ls_N).$$

З урахуванням усіх вищеописаних (підрозділи 2.1, 2.2) ознак P реферату R реферативна конструкція має такий загальний вигляд:

$$CK: Sr(Gr_8) - V(\bigvee_{i=1}^3 m_i \bigvee_{j=1}^2 Gr_j) - A_l(\bigvee_{i=4}^6 m_i (\bigvee_{j=1}^2 ls_j) Gr_3) - \bigvee_{l=2}^7 A_l(\bigvee_{i=4}^9 m_i (\bigvee_{j=1}^3 ls_j) (\bigvee_{r=4}^7 Gr_k))$$

Загальна модель реферату відповідно набуватиме такого вигляду:

$$R = \{CK_i(s_1), CK_i(s_2)\}, (i=1,2)$$

$$CK_1(s_1) : Sr(g_1^6) - \bigvee_{i=4}^5 (m_i g_1^4 g_2^1 g_3^2 g_4^2 g_8^2) - A_l(\bigvee_{j=1}^2 m_i \bigvee_{l=2}^7 ls_j g_1^1 g_7) - \bigvee_{i=4}^9 A_l(\bigvee_{j=1}^3 m_i (\bigvee_{k=4}^7 ls_j) (\bigvee_{k=4}^7 Gr_k)).$$

$$CK_1(s_2) : Sr(g_1^6) - \bigvee_{i=2}^3 (m_i g_1^4 g_2^1 g_3^2 g_4^2 g_8^2) - A_l(\bigvee_{i=4}^6 m_i \bigvee_{j=1}^2 ls_j g_1^1 g_7^1) - \bigvee_{l=2}^7 A_l(\bigvee_{i=4}^9 m_i (\bigvee_{j=1}^3 ls_j) (\bigvee_{k=4}^7 Gr_k)).$$

$$CK_2(s_1) : Sr(g_1^6) - \bigvee_{i=4}^5 (m_i g_1^3 g_2^2 g_3^2 g_4^1) - A_l(\bigvee_{j=1}^2 m_i \bigvee_{l=2}^7 ls_j g_1^1 g_7^1) - \bigvee_{l=2}^7 A_l(\bigvee_{i=4}^9 m_i (\bigvee_{j=1}^3 ls_j) (\bigvee_{k=4}^7 Gr_k)).$$

$$CK_2(s_2) : Sr(g_1^6) - \bigvee_{i=2}^3 (m_i g_1^3 g_2^2 g_3^2 g_4^1) - A_l(\bigvee_{i=4}^6 m_i \bigvee_{j=1}^2 ls_j g_1^1 g_7^1) - \bigvee_{l=2}^7 A_l(\bigvee_{i=4}^9 m_i (\bigvee_{j=1}^3 ls_j) (\bigvee_{k=4}^7 Gr_k)).$$

2. 3. Експериментальна перевірка роботи алгоритму автоматичного реферування

На основі побудованої моделі індикативного реферату й опису механізму заповнення актантної структури реферативних конструкцій було розроблено алгоритм автоматичного реферування.

Алгоритм складається з трьох етапів (рис. 2.6):

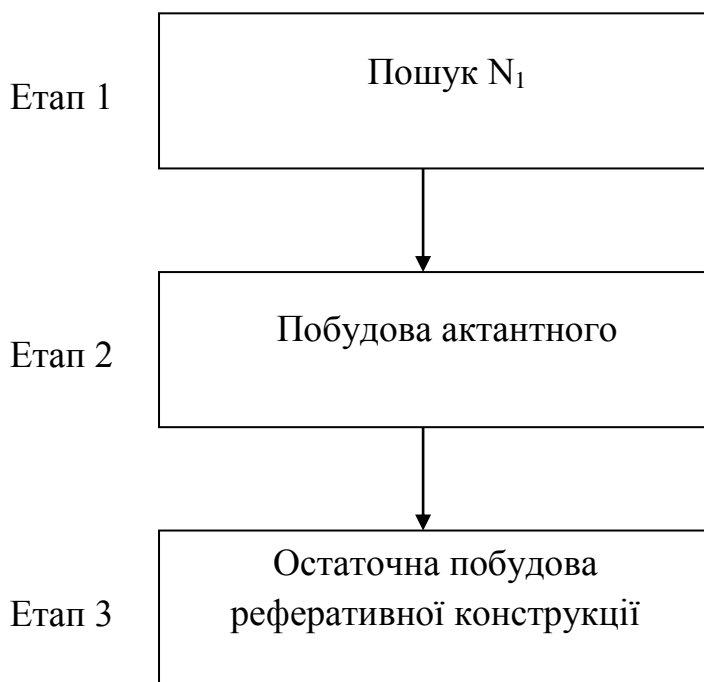


Рис. 2. 6 Основні етапи алгоритму автоматичного реферування

На першому етапі цього алгоритму відбувається пошук найчастіше уживаного поняття N_1 з тексту оригіналу для заповнення іменникового актанта A_1 в реферативній конструкції. Пошук здійснюється з опертям на заголовок первинного тексту, оскільки той обов'язково містить ключове термінологічне поняття. У процесі пошуку кожне слово заголовка порівнюється з кожним словом з оригінального тексту й у випадку збігу одночасно проводиться кількісний аналіз: підраховується абсолютна частота досліджуваних слів. У результаті відшукується найчастіше вживане поняття-термін N_1 , що передається на вхід другого етапу алгоритму реферування.

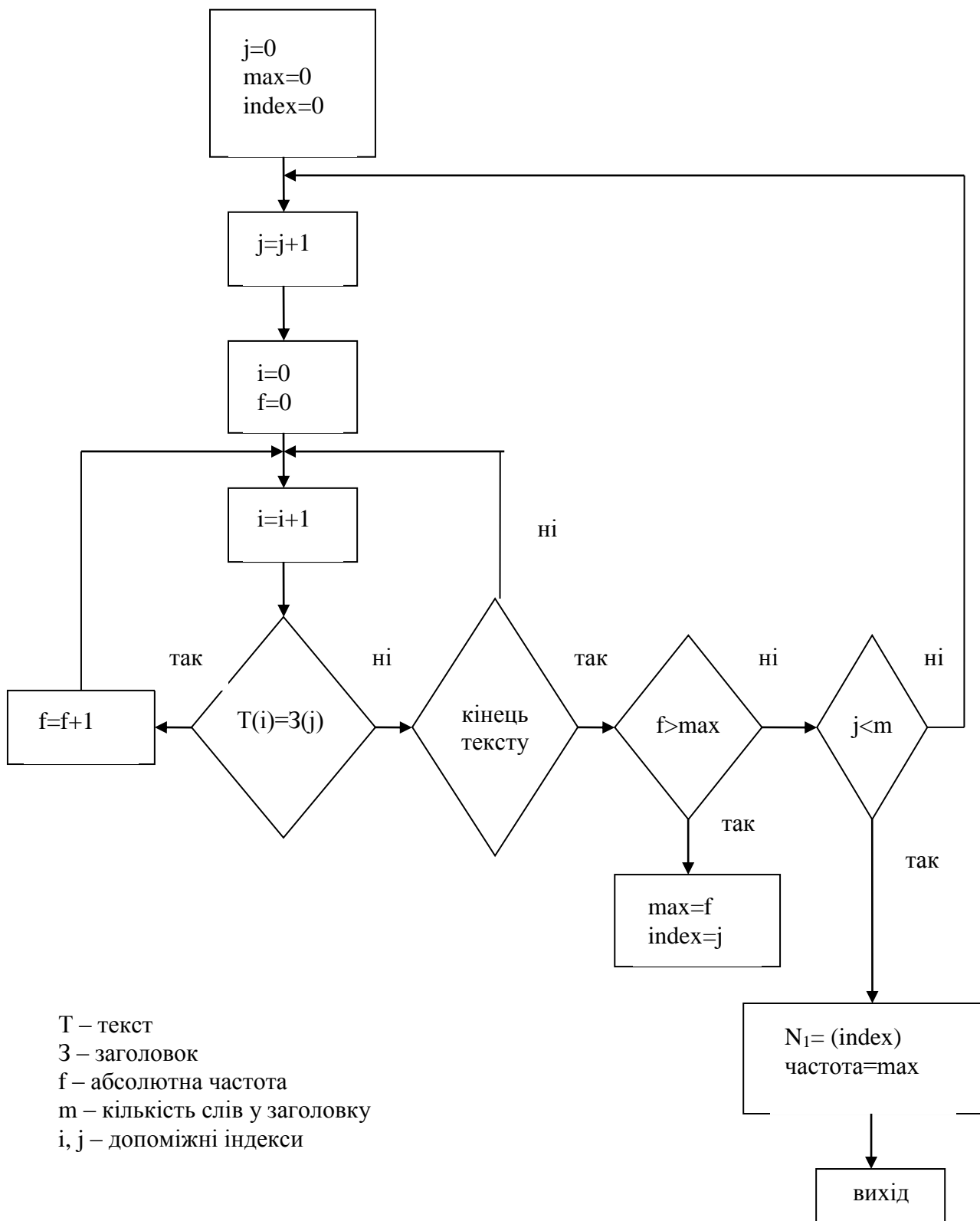


Рис. 2. 7 Етап 1 алгоритму реферування

Формування актантної структури $CK(s_1)$ розпочинається з заповнення іменникового актанта A_I терміном N_I із тексту оригіналу, що визначається

статистичним методом. Об'єкт або процес, описаний термінологічним поняттям, може бути самостійним об'єктом дослідження в тексті першоджерела, а отже, виступати в ролі A_1 в актантній структурі індикативного реферату. В переважній кількості випадків N_1 належить класові $N(m_4, ls_1)$, наприклад: *смысл, печень, текст, регион, транспорт, психика*.

Якщо прикметник *Adj* супроводжує іменник N_1 у більшості випадків його використання в оригінальному тексті або ж є окремою словниковою статтею у словнику предметної галузі, то $A_1=Adj N_1$: *железнодорожный транспорт, социальная психика, интегральный тип*.

Коли ж основне термінологічне поняття N_1 в тексті виражене поширеною іменниковою групою, воно використовується в скороченому вигляді, обов'язково позначеному в дужках після його першого повного формулювання. В рефераті віддається перевага використанню повної форми таких термінів, наприклад: *рак мочевого пузыря (РМП), словообразовательные значения (СЗ), язык запросов (ЯЗ), естественно-языковое высказывание (ЕЯВ)*... тощо.

Під час роботи автоматичної системи реферування на базі описаного алгоритму на усіх етапах проводиться автоматичний морфологічний аналіз із використанням SMART-технології.

На другому етапі алгоритму відбувається побудова актантного ланцюжка, який є семантичною основою реферативної конструкції. На вході в пам'яті утримується знайдений на першому етапі ключовий термін N_1 , який заповнює актант A_1 . Далі проводиться дистрибутивний аналіз цього елементу. Алгоритм для системи АР працює в двох блоках: умовно Блок 1 можна назвати *пошук вліво*, Блок 2 – *пошук вправо*, тобто в одному блоці аналізуються слова, які стоять зліва від N_1 , а в другому блоці – слова, які стоять справа. Таким чином, Блок 2 є своєрідним «дзеркальним» відображенням Блоку 1.

Перш за все аналізується слово N_2 , яке переважно при використанні терміна N_1 стоїть справа від нього. Якщо N_2 належить класам $N(m_4, ls_1)$ або

$N(m_9, ls_3)$, то N_2 заповнює актант A_2 , наприклад: *социетальная психика неслучайной группы, интегральный тип США, железнодорожный транспорт Украины.*

Далі розглядається лівостороннє слово, розташоване контактено з ключовим терміном у тексті оригіналу. Якщо N_3 належить класам $N(m_4, ls_1)$, $N(m_4, ls_2)$, $N(m_5, ls_1)$ або $N(m_5, ls_2)$, то слово N_3 заповнює актант A_1 , спричиняючи цим зсув індексів раніше заповнених актантів у бік збільшення на один:

$$A_1 = N_3$$

$$A_2 = N_1$$

$$A_3 = N_2$$

Наприклад: *математическое описание смысла, трансплантация печени, понимание текста, инвестиционная привлекательность региона, моделирование социетальной психики неслучайной группы, модель интегрального типа США, положение железнодорожного транспорта Украины.*

Згодом здійснюється статистичний аналіз (визначення частоти використання) слова N_4 , розташованого контактено зліва від слова N_3 у реченнях, що містять його. Якщо N_3 належить класам $N(m_4, ls_2)$ або $N(m_5, ls_2)$, то:

$$A_1 = N_4$$

$$A_2 = N_3$$

$$A_3 = N_1$$

$$A_4 = N_2$$

Наприклад: *необходимость математического описания смысла, аспекты трансплантации печени, проблема понимания текста, оценка инвестиционной привлекательности региона, улучшение положения железнодорожного транспорта Украины.*

Заповнення актантів A_i реферативної конструкції відбувається доти, поки в актантній структурі речень оригінального тексту не виявляться слова N_i , приналежні до класів $N(m_4, ls_2)$ чи $N(m_5, ls_2)$.

Актант A_i заповнюється словами зі значеннями «мета» й «засіб» відповідно у разі їх формального розпізнавання в тексті оригіналу, за умови, що головний елемент N_i належить класові $N(m_7, ls_2, g_1^1, g_7^2)$ або N_i належить класові $N(m_8, ls_2, g_1^1, g_7^3)$.

Актантна структура конструкцій $CK(s_2)$ відрізняється від $CK(s_1)$ наявністю загальнонаукового поняття зі значенням (m_6, ls_2) – властивість (оцінка). При побудові конструкцій $CK(s_2)$ початковий етап заповнення актанта A_1 відбувається за тією ж схемою, що й при побудові $CK(s_1)$. Однак за наявності в актантній структурі речень оригінального тексту слів зі значенням (m_6, ls_2) , алгоритм заповнення актантів змінює напрям, і відбір понять відбувається справа наліво у напрямку до N_1 : *актуальність изучения смысла, ефективность методов лечения рака мочевого пузыря*.

На третьому етапі завершується побудова реферативного речення. Оскільки на вході цього етапу вже є готовий актантний ланцюжок, який містить основний зміст остаточного речення для реферативного тексту, головним завданням цього етапу є вибір предиката зі словника $V(m_1)$ (підрозділ 2.2). Вибір предиката з першої групи, тобто зі словника $V(m_1)$ можна робити в автоматичному режимі, оскільки елементи цього класу є повністю взаємозамінними (підрозділі 2.2), крім того, наявність відібраного предиката в тексті оригіналу не обов'язкова.

Останнім етапом заповнення актантної структури речення індикативного реферату є заповнення актанта предикатного ядра V_i . Якщо в реченні, яке містить поняття N_1 в тексті оригіналу, функцію предиката виконує дієслово з граматичною ознакою $(g_1^3 g_2^1 g_3^1 g_4^3 g_5^1)$ та семантичною ознакою $(m_1 \vee m_2 \vee m_3)$ (*рассмотрим, проанализируем, сравним*), то предикатний актант реферативної СК заповнюється тією ж лексемою, але у формі $(g_1^4 g_2^1 g_3^2 g_4^2 g_8^2 \vee g_1^3 g_2^2 g_3^2 g_4^1)$. Однак за відсутності в тексті оригіналу таких лексем предикатний актант заповнюється лексемою з класу $V(m_1)$, $V(m_2)$ або $V(m_3)$ залежно від семантичного значення S речення CK . Так, предикатне ядро речення слугує

виключно для синтаксичної організації речення й виражене предикатом $V(m_1)$, наприклад:

Дано современное определение нестабильной стенокардии.

Изложена проблема обоснования и моделирования социетальной психики неслучайных групп.

Описывается структура языка запросов для системы управления базой данных.

Освещена проблема лечения рака мочевого пузыря.

Представлены основные аспекты пересадки печени.

Приводятся причины современной эпидемиологической ситуации по туберкулезу в мире и в Украине.

Рассматривается организация анализа естественно-языковых высказываний в диалоговых системах общения.

Рассмотрены вопросы усовершенствования финансово-кредитного механизма АПК.

Предикат речень $СК(s_2)$, презентуючи результат дослідження, описує його етапи і, як правило, виражений дієсловом $V(m_2)$, наприклад:

Доказана эффективность методов химиолучевого хирургического лечения, иммунотерапии.

Обоснована важность организации финансового консалтинга и развития лизинга сельскохозяйственной техники, оборудования и технологий.

Оценены актуальность и значимость дальнейшей перспективы трансплантации печени.

У випадку, коли предикат є носієм семантики авторської оцінки, воно виражене дієсловом $V(m_3)$, однак заповнення предиката дієсловом з означеного класу можливе лише за умови участі людини в побудові реферату. Граматичне значення предиката спричиняється типом синтаксичної конструкції, обраним для представлення інформації в тексті реферату.

При оформленні остаточного варіанта РК за допомогою предиката з другої і третьої груп діє зовсім інший механізм, тому що в цьому разі

необхідно, по-перше, знайти необхідне слово в первинному тексті, по-друге, проаналізувати його оточення для визначення рівня узагальнення і тільки тоді використовувати його для побудови РК.

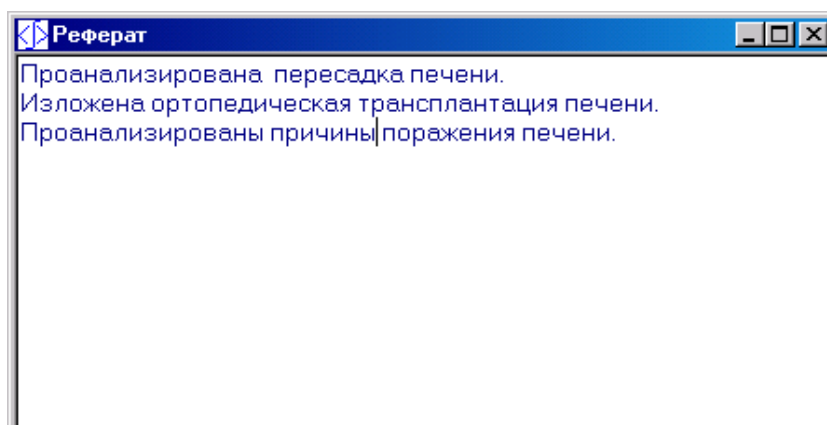
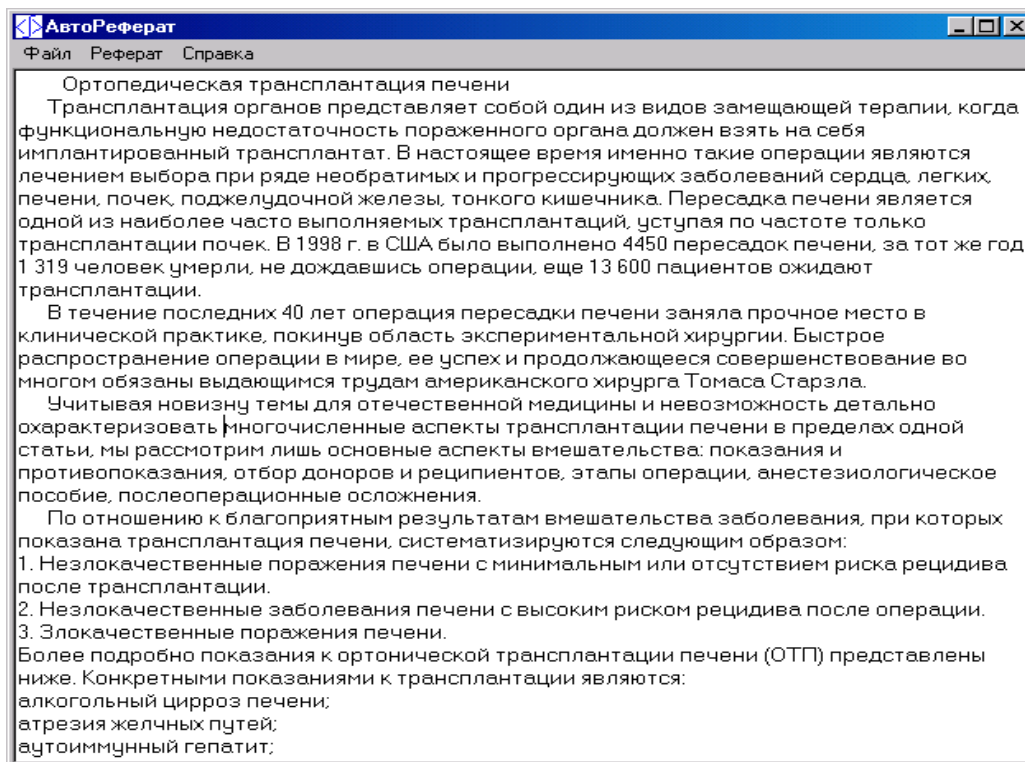
Для перевірки запропонованого алгоритму автоматичного реферування була створена експериментальна система «АвтоРеферат». Програмна реалізація здійснювалась на мові C++, середовище розробки – Borland C++ Builder 6.

«Автореферат» працює чітко за запропонованим алгоритмом. На першому етапі відшукується найчастіше уживане поняття шляхом зіставлення слів, які входять до складу заголовка, і слів оригінального тексту. Завдяки механізмам мов програмування, використанню додаткових змінних для зберігання даних здійснюється один перегляд первинного тексту. Іменниковий актант A_I на першому етапі заповнюється одним елементом – поняттям N_I , але на другому етапі до нього додаються можливі додаткові елементи, які складають термінологічне поняття. Наявність онтології предметної галузі на цій стадії має сприяти збільшенню якості й швидкості ідентифікації терміна.

На другому етапі аналіз оригінального тексту відбувається за блоками: спочатку пошук елементів актантного ланцюжка зліва, потім – справа. Для побудови ланцюжка здійснюється статистично-дистрибутивний аналіз контактних розташованих елементів (слів) первинного тексту. Критерієм відбору елементів у актантний ланцюжок слугує так званий коефіцієнт повторювань (k), який впливає на підрахунок абсолютної частоти.

На третьому етапі додається предикат методом випадкового вибору з допоміжного словника (виродженої онтології).

У процесі дослідження було проведено експериментальні перевірки роботи програми «АвтоРеферат» і їх порівняння з результатами реферування у Word. Експериментальна перевірка № 1 базувалася на текстах з медицини (рис. 2.8), а експериментальна перевірка № 2 відбувалася на тексті з математики (рис. 2.9).



K = 15%

Word

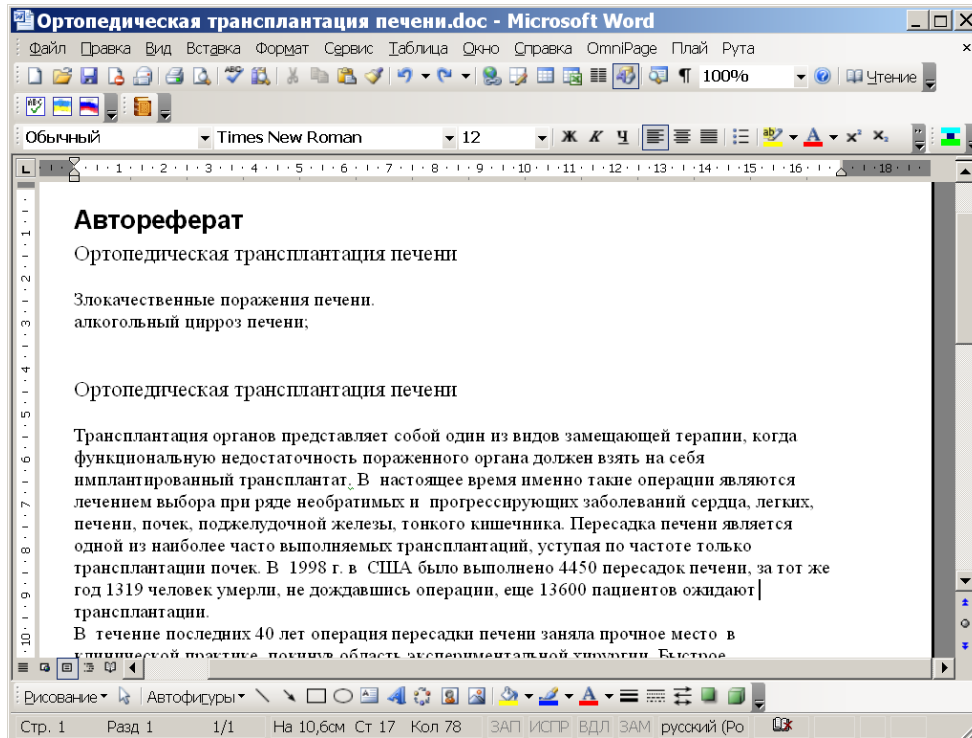
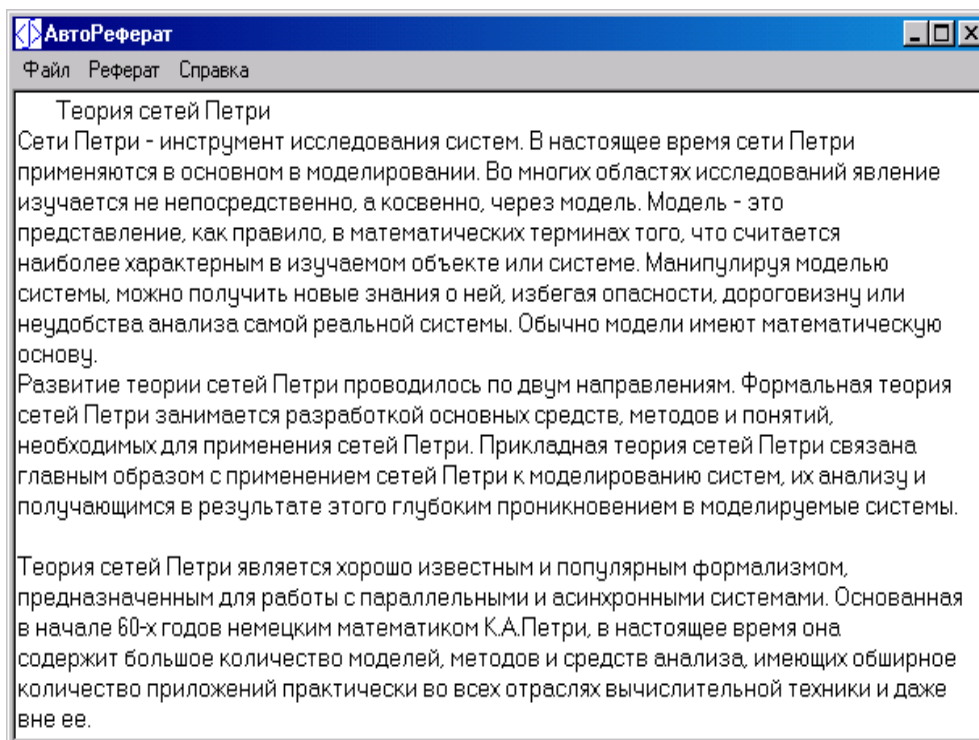
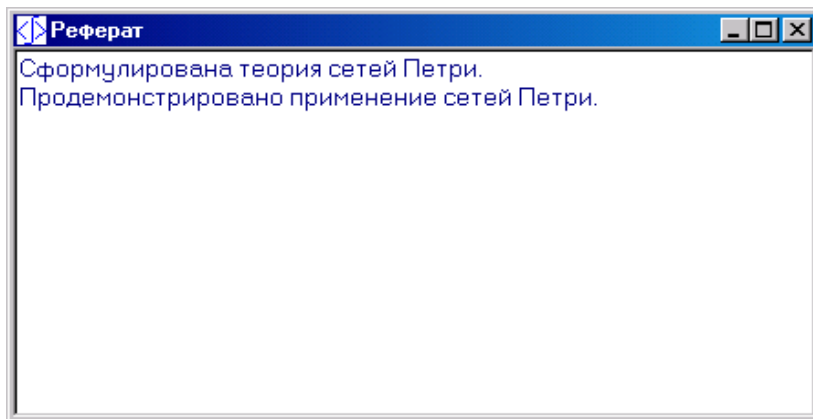


Рис. 2. 8 Результаты экспериментальной проверки № 1 (оригинальный текст и текст реферату)

№ 2





K = 20%

Word

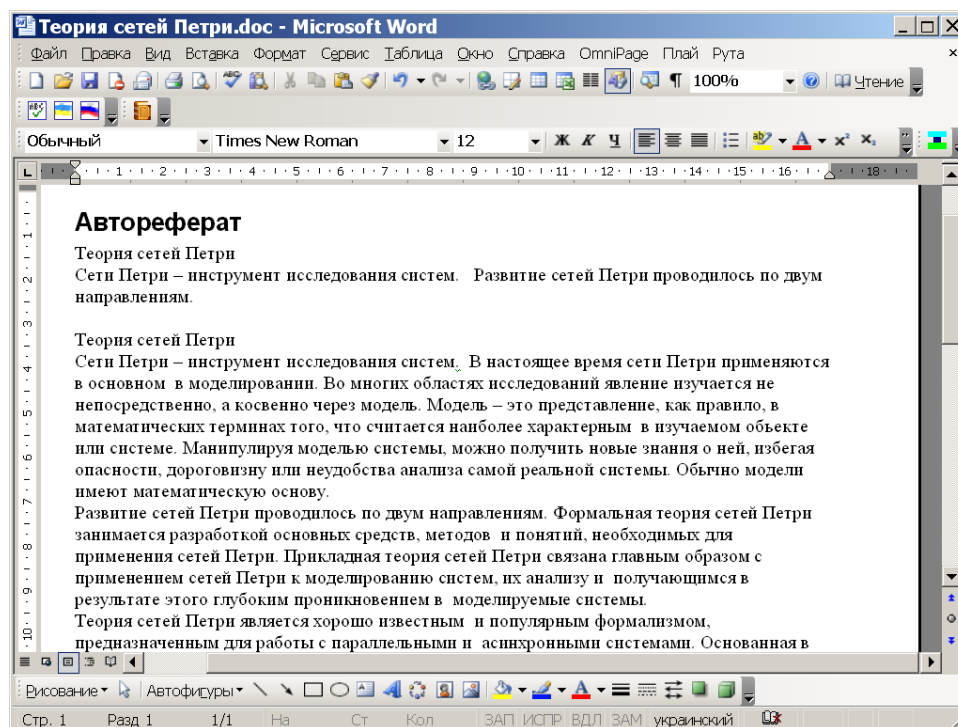


Рис. 2. 9 Результаты экспериментальной перевірки № 2 (оригінальний текст, текст реферату й автореферат у Word)

На виході програма «АвтоРеферат» дає реферативний текст, який формально відповідає побудованій у даному дослідженні моделі індикативного реферату, але ще не може генерувати повноцінний індикативний реферат. Кращі результати можливі лише після детального вивчення та опису семантичних відношень похідного тексту і його реферату.

Водночас порівняння результатів з авторефератами, які було отримано у Word, демонструє, що сформовані нами реферати являють собою тексти, котрі відповідають вимогам до індикативних рефератів як за змістом, так і за структурою, на відміну від рефератів, укладених з окремих речень тексту.

Описане в цьому розділі моделювання компресії тексту на всіх рівнях і побудована на базі проведеного аналізу структури реферату модель реферату дозволили розробити першу версію системи автоматичного реферування. Принаймні це лише тільки перший крок до майбутнього автоматичного реферату.

Висновки

1. Запропоновано підхід, у рамках якого об'єктами системи АР є модель індикативних рефератів і спеціальні засоби подання знань – текстова база, онтології предметних галузей та словник категорій реферативних конструкцій.

2. Формалізовано опис об'єкта дослідження – реферату, як тексту, що складається з набору речень певного типу.

3. Побудовано модель реферату у вигляді набору типових для індикативних рефератів синтактико-семантичних конструкцій (структур). Типовий індикативний реферат складається з одного – трьох речень, які представляють собою варіанти поширення типових синтаксичних конструкцій.

4. Реферат описується набором інваріантних ознак, що характеризують синтаксичні, семантичні і, частково, граматичні зв'язки реферату з вихідним текстом.

5. Модель компресії на синтаксичному рівні представляє собою дві типові реферативні синтаксичні конструкції із семантикою результативності.

6. Описано граматичні значення актантів і сирконстантів синтаксичних конструкцій.

7. Модель компресії на семантичному рівні містить у собі три складові:

– *семантико-синтаксичну*, яка описує правила побудови семантичної структури реферативних речень у вигляді набору елементів: предиката, об'єкта, що є граматичним суб'єктом, актантів і сирконстантів;

– *власне семантичну*, яка описує правила формування семантичних значень речень у складі реферату у вигляді набору речень із семантичним значенням об'єкта або результату дослідження;

– *лексико-семантичну*, яка описує правила класифікації лексем, що беруть участь у заповненні актантних структур реферативних конструкцій.

8. Розглянуто заголовок як реферат мінімального обсягу або як текст із максимальним рівнем згортання змісту.

9. Описано смислові структури словосполучень на різних рівнях згортання інформації, що дозволило побудувати низку переходів від заголовка до тексту і далі до реферату під час його змістового конструювання.

10. На основі побудованої моделі індикативного реферату й детального опису механізму заповнення актантної структури реферативних конструкцій було розроблено алгоритм автоматичного реферування з використанням на даному етапі головним чином статистичних методів аналізу тексту.

11. Створено програму «АвтоРеферат», яка породжує реферативний текст, що формально відповідає побудованій у даному дослідженні моделі індикативного реферату, але ще не дає можливості генерувати повноцінний індикативний реферат. Кращі результати можливі лише після детального вивчення й опису семантичних відношень похідного тексту та його реферату.

12. Порівняно з авторефератами у Word сформовані реферати являють собою тексти, що відповідають вимогам до індикативних рефератів як за змістом, так і за структурою, на відміну від авторефератів, укладених з окремих речень тексту.

III. МОДЕЛЮВАННЯ СЕМАНТИЧНИХ ЗВ'ЯЗКІВ «ТЕКСТ-РЕФЕРАТ» ІЗ ВИКОРИСТАННЯМ ОНТОЛОГІЙ

3. 1. Побудова семантико-контекстної моделі реферування

На сьогодні для розробки систем автоматичної обробки текстів широко використовуються технології семантичного аналізу, що ґрунтуються на формуванні формального опису змістового навантаження тексту у вигляді фреймів, семантичної мережі або інших способів представлення знань. Одним із шляхів у розв'язанні цього завдання є побудова когнітивної або семантико-контекстної моделі, яка забезпечила б глибоке проникнення у зміст тексту і його трансформацію зі збереженням смислу.

У нашому дослідженні для формального опису змісту тексту й формального опису смислових перетворень у процесі реферування використовуються:

1. заголовок, який презентує зміст вихідного тексту в концентрованому вигляді;
2. текстова база, яка містить речення, що є «інформаційним ядром» тексту, котре утримує інформацію, залежну від ситуації (тематики тексту);
3. онтології, що містять незалежну від тематики тексту інформацію: онтології верхнього рівня (набір смислових категорій, які описують смислові аспекти, котрі входять до реферату), онтології загальнонаукової лексики й онтології предметних галузей.

Вивчення впливу кожного з них на аналіз змісту тексту з метою його реферування стало завданням другого етапу наших досліджень.

На нашу думку, аналіз змісту тексту повинен містити аналіз заголовка, тому що заголовки науково-технічних статей дають уявлення про основний напрямок змісту статті. Важлива риса заголовка – кількість значущих слів у ньому, тобто слів, які несуть основне змістове навантаження і є потенційно ключовими. При цьому заголовки, звичайно, пишуться максимально стисло, надто лаконічно, в них опущені всі семантично другорядні елементи. Отже,

йдеться про компресію змісту тексту. Більше того, ми розглядаємо заголовок як реферат мінімального обсягу або як текст із максимальним рівнем згортання змісту.

У зв'язку з цим був проведений порівняльний аналіз компресії в заголовку й рефераті (рис. 3.1).

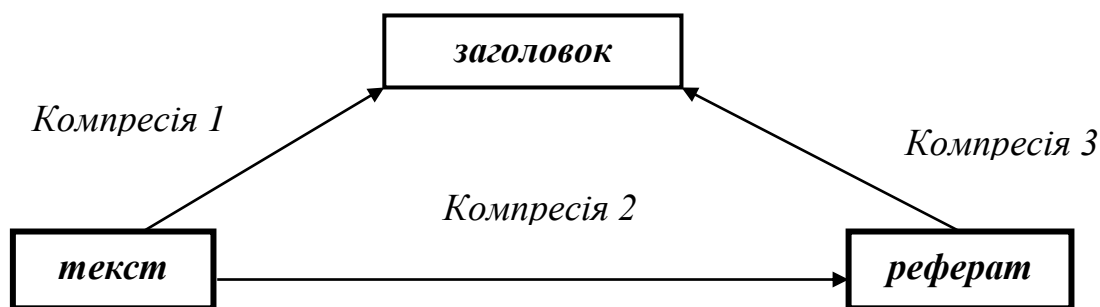


Рис. 3.1 Згортання інформації

У результаті дослідження змістової і синтаксичної структури заголовка виявилася його схожість зі структурою реферату. Як і в індикативному рефераті змістова структура заголовка складається з двох метазначень – «об'єкт» і «результат». Утім, на відміну від реферату, вони є елементами змістової структури одного речення і вживаються у зворотній, порівняно зі змістовою структурою реферату (*об'єкт – результат*), послідовності: *результат – об'єкт*. Така подібність змістових структур реферату й заголовка, що містять однакові змістові аспекти, стала підставою для вивчення взаємозв'язку текстів і заголовків, аби за допомогою інформації, що міститься в заголовку, виявити в тексті ті лексичні одиниці, які необхідні для семантичного наповнення моделі реферату даного тексту.

Для цього було побудовано класифікацію лексем СК заголовка [102]. Усі дієслова-предикати розбито на три класи зі значенням: m_1 – форми подання інформації в тексті, m_2 – етап роботи, m_3 – порівняльна оцінка. Лексичні значення іменників, що входять до складу іменникових груп, поділено на шість класів зі значенням: m_4 – об'єкт, m_5 – результат (або процес, що прагне досягнення результату), m_6 – ознака, m_7 – мета, m_8 – засіб і m_9 – місце.

Слід зауважити, що при описі заголовка використовується спільна з описом реферату система позначень, оскільки обидві є частинами розробки спільної моделі процесу реферування. Тому класи віддієслівних іменників визначаються при описі заголовків як клас $V(m_5)$, що відповідає класу дієслів, які виступають у ролі предикатів реферативних конструкцій. Усі інші визначаються як класи $A(m_i)$ з відповідними метазначеннями m_i .

Слова, які належать до класів зі значеннями m_4 і m_5 , є основними компонентами структури заголовка. При цьому, на відміну від реферату в заголовковій конструкції на першому місці, як правило, стоїть актант «результат» (процес, що прагне досягнення результату).

$СК_3 : V(m_5) - A(m_4)$

Наприклад: *О развитии социума.*

Іноді, дуже нечасто, актант «результат» може бути відсутній узагалі.

$СК_3 : A(m_4)$

Наприклад: *Рынки, правительства и глобальное руководство.*

Відмінна риса заголовків полягає в тому, що практично в усіх заголовках відсутній предикат у чистому вигляді. Він, зазвичай, трансформується у віддієслівний іменник чи іменник зі значенням дії. Це відповідає відомому у стилістиці факту: заміна віддієслівних іменників дієсловами сприяє конкретизації тексту. Звідси випливає: трансформація дієслова підвищує абстрактність, а отже, й узагальненість тексту.

Якщо розглядати класи зі значеннями m_1 , m_2 і m_3 як віддієслівні іменники, то можна сказати, що m_1 як форма подання інформації в тексті (*изложение, освещение, описание, введение, рассмотрение*) зазвичай не наявна в заголовку, m_3 демонструє суб'єктивну оцінку автора (*дополнение, изменение, исправление, улучшение*), що також зустрічається вкрай рідко, а m_2 – трансформується в m_5 , яка виступає в ролі результату і є одним із двох основних компонентів структури заголовка. Наприклад, у рефераті: *Анализируется инвестиционная привлекательность различных регионов*

України. А в заголовку: *Анализ инвестиционной привлекательности различных регионов Украины.*

Клас $V(m_5)$, де m_5 – етап роботи, містить у собі лексеми, які мають метатекстові значення – опис визначених етапів проведеної роботи [208].

При аналізі заголовків з'ясувалося, що деякі лексеми, утворені від предикатів, збігаються з лексемами, що відносяться до класу $A(m_2)$:

$V(m_5) = \{ \text{анализ, использование, исследование, обобщение, оценка, реализация, решение, установление} \}$. Сюди також можна додати такі лексеми, що не зустрічалися в дослідженні: $\{ \text{взаимодействие, восстановление, идентификация, моделирование, обращение, определение, организация, планирование, построение, проектирование, развитие, реструктуризация, синтез, формализация} \}$.

Інші компоненти структури заголовка можуть бути присутні, а можуть і не зустрічатися в певних заголовках. Усе залежить від того, наскільки важливим є цей компонент у тексті. За результатами порівняльного аналізу з текстом реферату встановлено, що всі компоненти структури заголовку відповідають компонентам структури реферату.

При вивченні граматичних характеристик актантів заголовкових структур було проведене порівняння з рефератом: переважно всі поняття з метазначенням «результат» у заголовку стоять у називному відмінку. Однак є ряд заголовків, коли перед актантом «результат» стоїть прийменник (*о/об, про, к*) чи указник на результат з прийменником або без нього (*к вопросу о, гипотеза о, учителя о, еще раз о, к заданию, одним из вопросов*). Залежно від цього наступні за ними лексеми, що напевне виражають *об'єкт* чи *результат*, можуть виступати в трьох відмінках: родовому (*К проблеме управления (руководства) вузовским обществом*, давальному (*К взаимодействию религиозного и советского образования*) і місцевому (*О развитии социума*).

Поняття з метазначенням «об'єкт» найчастіше виступає у формі іменника в родовому або називному відмінку за умови, що перед ним не стоїть актант «результат».

Поняття з метазначенням «мета» вживається тільки у формі родового відмінка з прийменником (*Процедура дедуктивного вивода для планирования работы системы управления в динамической среде*), поняття з метазначенням «засіб» – у формі орудного (*Об алгоритме исправления орфографических ошибок методом корректирующего словаря*), а поняття з метазначенням «місце» – місцевого відмінка (*Обзор исследований по автоматической обработке естественного языка в США*), що цілком збігається зі структурою реферату.

На основі проведеного дослідження була побудована загальна модель заголовка:

$$СК_3 : O(\bar{b}) / K - Sr - V(m_5) - A(m_4) - A(m_7) - A(m_9) - A(m_8).$$

Жирним шрифтом виділено обов'язковий елемент заголовка, наявність усіх інших – можлива.

Модель складається з обов'язкових елементів – актантів $A(m_i)$ і необов'язкових елементів – сирконстантів Sr . Прислівники, що заповнюють сирконстанти, є носіями оцінювальної семантики, тому заповнення сирконстантів є винятково інтелектуальним процесом (тобто може виконуватися лише людиною). Актант «об'єкт» у даній структурі є основним її компонентом, актант «результат» часто – але не завжди – наявний, інші компоненти, як зазначалося, наявні залежно від їхньої значущості в тексті, якому відповідає заголовок.

У повному вигляді дана модель зустрічається рідко, оскільки містить у собі всі, в принципі можливі, варіанти. Зазвичай у ній наявні від двох до чотирьох компонентів:

$$СК_3 : V(m_5) - A(m_4) - A(m_7) - A(m_9).$$

Процедура дедуктивного вивода для планирования работы системы управления в динамической среде.

$$СК_3 : O - V(m_5) - A(m_4) - A(m_7).$$

О формировании семантических признаков для математической модели префиксального словообразования.

$CK_3: O - Sr - V(m_5) - A(m_4) - A(m_9).$

Об одной реализации языка запросов в системе управления базой данных.

Проведені дослідження дозволили відстежити чітко продовження процедури згортання (компресії) в заголовку перед процедурою компресії в рефераті як на семантичному, так і на синтаксичному рівнях. При збереженні тих самих змістових складових у рефераті й у заголовку вони мають також синтаксичні й граматичні особливості їх вираження, які підвищують рівень компресії тексту заголовка.

Наявність одних і тих самих семантичних компонентів у заголовку, рефераті й тексті, що є різними формами вираження одного й того ж поняття, дозволяє описати смислові структури словосполучень на різних рівнях згортання інформації, для чого необхідна наявність онтологій предметних галузей та загальнонаукової лексики. І лише тоді можна побудувати низку переходів по цих структурах від заголовка до тексту й далі до реферату при його змістовому конструюванні. Для здійснення переходу від заголовка до тексту й далі до реферату, необхідно побудувати текстову базу, до якої входять речення, що містять слова з заголовка або ж їх смислові еквіваленти з тексту.

Таким чином, ще одним необхідним елементом у роботі системи АР є текстова база знань. Вона складається з фактів і тверджень, пов'язаних із певною ситуацією (конкретним текстом). І на противагу онтології, яка містить незалежну від ситуації і стану інформацію, є «інформаційним ядром», що містить залежну від ситуації і стану інформацію.

Для побудови текстової бази знань на цьому етапі досліджень ми відштовхувалися від понять, які містяться в заголовку документа. За ключовими словами, знайденими в заголовку відшукуються відповідні їм іменникові групи в тексті (будується текстова база знань), після чого формуються ланцюжки іменникових груп для реферативних конструкцій відповідно до наявної моделі реферату.

Текстова база мусить містити інформацію, виражену власне текстом, тобто референціальне значення тексту, оскільки, як відомо, на відміну від художньої літератури для наукової і технічної літератури характерне превалювання референціальних значень.

Згідно з концепцією, запропонованою в роботі [60], для опису глобального змісту тексту необхідна побудова схеми, що забезпечувала б «швидкий аналіз поверхових структур і побудову простої та жорсткої семантичної конфігурації». Відповідно до даної концепції, розуміння тексту припускає часткове планування (або очікування) структур і значень речень і цілого тексту. Такі структури тексту (називаються макроструктурами) являють собою узагальнений опис основного змісту дискурсу, який читач буде в процесі розуміння, і вважаються фактично рефератом або резюме.

Побудова макроструктури здійснюється в процесі застосування макроправил. До ряду таких правил належать правила скорочення (несуттєвої інформації), узагальнення (двох чи більше однотипних пропозицій) і побудова (тобто комбінації кількох пропозицій в одну), що, власне, є еквівалентним побудованим нами моделям компресії тексту й моделі реферату.

Передбачається, що послідовно застосовуючи макроправила, можна побудувати формальний перехід від вихідного тексту до реферату, який складається з кількох речень. Побудова макроструктур читцями – це один із різновидів так званих стратегій розуміння дискурсу.

Поряд із цим схема дозволяє «забезпечувати певні загальні обмеження на можливі локальні й глобальні значення текстової бази», що відповідає обмеженню на змістові аспекти, які ми виділяємо в тексті для змістового наповнення моделі реферату.

Оскільки «процес розуміння передбачає часткове планування (або очікування) (у нашому випадку очікування) «структур і значень речень та цілих текстів», то саме побудова текстової бази (ТБ), що являє собою основу макроструктури тексту, дозволить наблизитися до формального опису деяких аспектів процесу розуміння тексту людиною [60].

Відштовхуючись від цих міркувань і розроблених нами на першому етапі дослідження моделей різноманітних етапів реферування, ми змогли підійти до розв'язання завдання побудови семантико-контекстної моделі реферування, яка може сприяти глибинному проникненню в зміст текстів і його трансформації зі збереженням сенсу.

Із цією метою ми використали виявлену й описану схожість змістових структур реферату й заголовка для вивчення взаємозв'язку текстів і їх заголовків для того, щоб за допомогою інформації, яка міститься в заголовку, виявити в тексті ті лексичні одиниці, що є необхідними для семантичного наповнення моделі реферату екстенсіональною семантикою даного тексту.

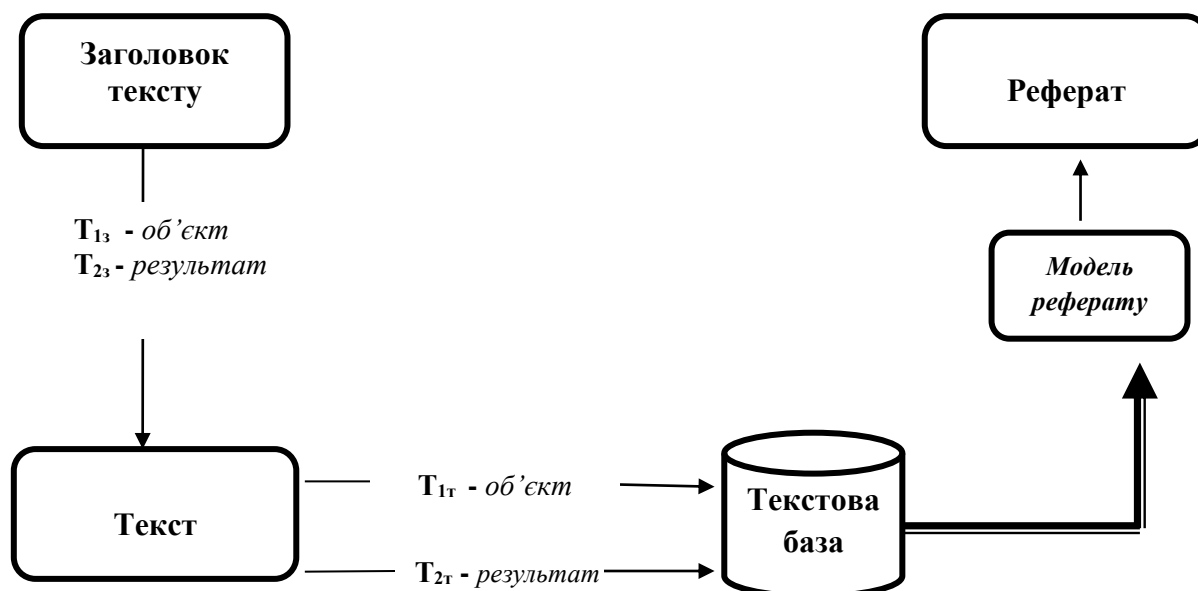


Рис. 3. 2 Схеми реферування на основі текстової бази

У процесі дослідження наукових текстів ми проаналізували змістову й структурну зв'язність наукового тексту. Перші результати наших досліджень показали, що змістова зв'язність заголовка, який виступає в ролі своєрідного ключа до розуміння тексту, й безпосередньо тексту, здійснюється переважно за допомогою засобів лексемрекурентності, імпліцитної рекурентності й референційної рекурентності. Використання того чи іншого типу зв'язності визначається структурно-композиційною характеристикою тексту.

На сьогодні ведеться робота над побудовою схеми, яка забезпечує швидкий аналіз поверхневих структур тексту за рахунок використання слів-

указівників на змістові аспекти в тексті, необхідних для побудови реферату (об'єкт, результат, мета, засіб).

Однак не завжди виділення речень за допомогою слів-указівників дозволяє здійснити оптимальний вибір речень із тексту для текстової бази. Для того щоб бути впевненим у правильності вибраних речень, необхідна наявність у системі автоматичного реферування онтологій, у яких достеменно зафіксовані всі концепти відповідної предметної галузі.

Водночас саме поняття онтології не є сьогодні однозначно інтерпретованим. Для вибору інтерпретації цього поняття, яка відповідала б нашим завданням, було проведено аналіз чинних на сьогодні підходів до визначення й варіантів використання онтологій.

3. 2. Обґрунтування вибору онтологій як засобу подання знань у системі автоматичного реферування та мови їх опису

Необхідність використання в системі автоматичного реферування онтологій, як одного з найбільш перспективних способів подання знань, стала предметом нашого дослідження з огляду на неоднозначність інтерпретації цього поняття в сучасній науці. З цією метою було здійснено аналіз чинних визначень онтології і обрано те, що відповідає завданням аналізу змісту тексту й подальшого подання його у вигляді реферату, яким ми розуміємо його сьогодні. Також було розглянуто варіанти використання онтологій у сучасних інформаційних системах.

Термін «онтологія» існує в науці тривалий час і зазнав чимало змін в інтерпретації. У філософському контексті онтологія розглядається як певна система категорій, які є наслідком певного погляду на світ [41, ст. 284]. Онтологія (з грецької *онтос* – сутність, *логос* – учення, поняття) – визначає вчення про буття, про сутність, його основи, принципи, структуру й закономірності (на відміну від гносеології – вчення про пізнання). Вже у Х. Вольфа (1679–1754), автора терміна «онтологія», вчення про буття відокремлене від учення про пізнання. У філософську літературу його укорінено німецьким філософом Р. Гокленіусом (1547–1628). На той час онтологія була складовою метафізики, наукою самостійною, незалежною і не пов'язаною з логікою, «практичною філософією», науками про природу. Її предмет становив вивчення абстрактних і загальних філософських категорій – *буття, субстанція, причина, дія, явище* тощо, а сама онтологія як наука претендувала на цілковите пояснення причин усіх явищ.

Але на початку 90-х років ХХ ст. поняття «онтологія» було переосмислено на ниві досліджень зі штучного інтелекту. В загальному значенні під онтологією тут розуміють опис формальною мовою понять певної предметної галузі й відносин між ними [78, ст. 28]. Онтології накладають певну структуру на ПГ й обмежують можливі інтерпретації термінів. Вони багато в чому схожі на тезауруси й таксономії, але насправді ширші за них,

оскільки представляють додаткові засоби для опису структури даних. Загальноприйнятим і найбільш поширеним сьогодні вважається визначення онтології, зроблене Т. Грубером: «онтологія – це формальна, точна специфікація концептуалізації» [232, ст. 907–928]. Існує принципова відмінність між мовою для вираження знань і знаннями, вираженими цією мовою. Так, концептуалізація, як абстрактна модель, не є залежною від мови. Онтологія ж, яка представляє дану концептуалізацію, посиляється на поняття, визначені словником мови, і, отже, залежна від уживаної мови.

Звідси вважається, що онтологія формально складається з організованих у таксономію термінів, як загальних, так і специфічних для конкретної ПГ, їхніх визначень і атрибутів, а також пов'язаних з ними аксіом і правил висновку. Онтологія дійсно завжди є проявом розуміння ПГ. Однак у кожній галузі може існувати неоднозначне розуміння тих самих термінів: «онтологія є консенсус відносно ПГ для певних цілей» [284]. У цьому випадку онтологія застосовується для структурування інформації, будучи посередником поміж людино- і машинно орієнтованим рівнем подання інформації. Запровадження онтологій дозволяє спростити процес спільного й багаторазового використання знань.

Залежно від сфери застосування (починаючи від інженерії знань, інформаційних і лінгвістичних теорій, закінчуючи логічними теоріями) поняття онтології суттєво варіюється у зв'язку з необхідністю використання її для розв'язання різноманітних завдань [78, 210, 227, 230, 244, 269, 283, 290, 291, 297].

Ми погоджуємося з точкою зору авторів, за якою онтологія – це експліцитна специфікація концептуалізації, де в ролі концептуалізації виступає опис значної кількості об'єктів предметної галузі і зв'язків між ними [232]. І вважаємо, що модель онтології кожної ПГ повинна містити як формальні елементи, так і їх змістове тлумачення в термінах, зрозумілих фахівцям цієї ПГ, тобто поняття, визначені в словнику, повинні бути прийнятими в даній ПГ галузі термінології [84].

З погляду концептуалізації, що описується, залежно від конкретного завдання або ПГ онтології поділяються на такі [41, ст. 299]:

– *онтології подання знань*, що визначають концептуалізацію, яка складає основу формалізму подання знань. Поняття і відносини, визначені в інших видах онтологій, вважаються конкретизацією понять онтології даного виду. Вони не залежать від конкретної ПГ і можуть бути використані для підтримки декількох альтернативних теорій, причетних до однієї і тієї ж ПГ;

– *загальні (родові) онтології*, що зачіпають загальні, фундаментальні аспекти концептуалізації – «частина», «причина», «участь», «подання» тощо;

– *проміжні онтології*, що містять загальні поняття і відносини, характерні для конкретної ПГ. В ідеальному випадку, вони запроваджуються як інтерфейс між онтологіями ПГ і загальних онтологій, але можуть виступати, як онтології верхнього рівня для опису знань ПГ;

– *онтології верхнього рівня*, що є конкретним призначенням понять загальних і проміжних онтологій. Цей модуль описує загальні поняття, які не залежать від конкретної проблеми або галузі [254];

– *онтології предметних галузей*, що містять поняття певної галузі знань або галузей, щодо неї входять [288];

– *онтології завдань*, що описують певні завдання або діяльність, із використанням словника, введеного в загальних, проміжних онтологіях і онтологіях ПГ [292];

– *онтології-додатка*, що є спеціалізацією онтологій ПГ й онтологій завдань, які ґрунтуються на визначеннях, притаманних конкретному додатку [292].

Наведений список видів онтологій не можна назвати повним і несуперечливим. Водночас він охоплює головні різновиди онтологій, які найчастіше використовуються в сучасних інформаційних системах. Для розроблюваної нами системи реферування ми спочатку взяли за формування словників для трьох типів онтологій:

– *онтології верхнього рівня*, обмеживши її обсяг на даному етапі досліджень категоріями, які використовуються в індикативних рефератах (об'єкт, результат, мета, засіб) і водночас визначають місце реферату в загальному наборі категорій верхнього рівня;

– *проміжної онтології загальнонаукової лексики*, яка дозволить пов'язати загальнонаукові поняття з поняттями різноманітних предметних галузей;

– *онтології предметних галузей* (у нашому дослідженні планується розробка онтології лінгвістичної термінології).

Процес розробки онтології зазвичай починається з укладання словника термінів (понять), які надалі будуть використаними для дослідження їхніх властивостей і характеристик.

Потім природною мовою створюється список точних визначень термінів, що містяться у словнику. Тоді на основі таксономічних відносин будуються дерева класифікації понять (ієрархії класів), яких у онтології може бути декілька.

Із понять, не задіяних в укладанні дерев класифікації, виділяються атрибути класів і їх можливі значення. Саме ці поняття й установлюють основні зв'язки між класами, в результаті чого виходить модель системи.

І на останньому етапі експерти з ПГ формують правила логічних висновків, які дозволяють оперувати поняттями й даними, а також здобувати зі створеної онтології нові знання [78, ст. 30].

На даному етапі наших досліджень укладений словник загальнонаукової лексики являє собою вироджену онтологію (Додаток Г), яка слугує основою для побудови онтології загальнонаукових термінів.

Згідно з розробленою моделлю реферату укладено словник реферативних категорій, які описують змістові аспекти, що формують змістову структуру індикативного реферату і являють собою фрагмент нижнього рівня категорій *онтології верхнього рівня*, який, відповідно, описує загальні поняття, незалежні від конкретної проблеми або галузі.

Прикладом досить повної онтології верхнього рівня є онтологія, наведена на рис. 3. 3.

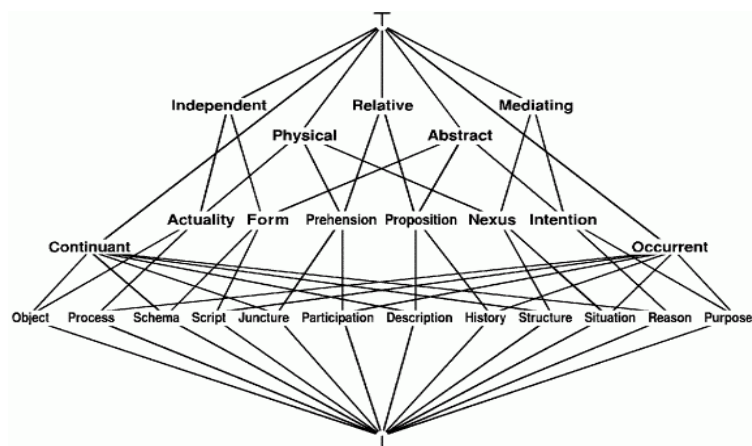


Рис. 3. 3 Онтологія верхнього рівня

На сьогодні більша частина проектів, які тією чи іншою мірою пов'язані з онтологіями, наявна в Інтернеті [225; 232; 236; 237; 260; 284; 290; 291]. Істотний інтерес для нас становлять ті, що демонструють можливості використання онтологій у системах автоматичного реферування.

Головна мета використання онтологій в Інтернеті міститься в перетворенні Інтернету як бази знань, на систему, ґрунтовану на знаннях, що намагалися здійснити розробники проекту Knowledge Acquisition Initiative (KA)² [212]. (KA)² – це відкрита ініціатива, де учасники долучаються до процесу створення розподіленої онтології і моделі здобуття знань (онтології ПГ). Одним з етапів організації інтелектуального доступу до документів у цьому проекті є реферування Web-документів, що здійснюється за допомогою онтологій.

Більшість систем автоматичної обробки інформації в мережі Інтернет, орієнтованих на підвищення якості обробки через інтелектуалізацію цього процесу, використовують реферування як необхідну його складову. Так у проекті SHOE (Simple HTML Ontology Extensions) запропоновано реферувати інформацію, яка міститься в HTML-документах, через онтології [253]. Дякуючи цьому з'явиться можливість вносити до Web-сторінки знання, доступні інтелектуальним агентам для читання HTML, збагаченої низкою спеціальних тегів. Це дозволяє знаходити знання за допомогою таксономій і

правил висновку, що існують в онтології. Завдяки цьому надається більш релевантна відповідь на запит перед наявними пошуковими машинами.

В SHOЕ користувачі в змозі самостійно реферувати свої документи й поширювати онтології новими поняттями, додаючи до них, таким чином, семантичний зміст. Але оскільки в цій системі центральний адміністратор онтологій не визначений, користувач, здійснюючи запити, може не знати всіх термінів, що використовуються для анотації HTML-документів. Тому у відповіді на запит траплятиметься брак важливої для нього інформації.

У проекті Ontobroker також запроваджена організація онтології автоматичного реферування Web-документів [224]. Визначається група фахівців зі спільним поглядом на поняття і їх місце в загальній онтології. На противагу SHOЕ уведено адміністратора онтологій; що ж до користувачів, то вони в змозі розпізнати всі терміни онтологій. Якщо SHOЕ ґрунтується на дескриптивній логіці як базисному формалізму для висновку, то Ontobroker започатковує логіку, засновану на фреймах, і має доволі складний механізм виведення відповіді на запити. Створений у межах цього проекту спеціальний пошуковий механізм Ontocrawler підтримує повну колекцію всіх реферованих HTML-сторінок.

Існують окремі завдання, які потребують наявності модуля реферування. Прикладом цьому слугує проект Plinus, метою якого є напівавтоматичне вилучення знань із текстів щодо механічних властивостей керамічних матеріалів природною мовою. Через те що тексти представляють широкий діапазон понять, у проекті використано багато інтегрованих онтологій для охоплення таких понять, як керамічні матеріали і їхні властивості, способи обробки, різні дефекти.

Окрім розв'язання окремих завдань (пошуку інформації, власне реферування та ін.) існують проекти, орієнтовані на розробку баз знань для роботи будь-яких систем, ґрунтованих на знаннях. Одним із таких проектів вважається СУС[®]. Проектом СУС[®] створена мультиконтекстна база знань і машини виведення (розроблена Cусогр). Основна мета проекту – побудувати

доступну для різноманітних програмних засобів базу знань усіх загальних понять, яка б обіймала семантичну структуру термінів і відношень між ними. [254].

Підсумовуючи наведений вище огляд перспектив розвитку онтологій, наголосимо, що цей вид систем суттєво розширює сферу застосовування інтелектуальних технологій (через підвищення їхньої якості). Дотепер, як зазначає Н. В. Майкевич, із виникненням баз знань і систем, заснованих на знаннях, з'являються нові перспективи, зокрема, при опрацюванні текстової інформації [113].

Навіть стислий аналіз стану розробок у сфері створення інтелектуальних систем переконливо свідчить, що проблема ще далека від свого остаточного вирішення, натомість вона є вкрай актуальною і потребує нових підходів і ідей, подальшого опрацювання. Звичайно, існує ряд дискусійних моментів, оскільки навряд чи можна очікувати на повсюдне залучення онтологій пересічним користувачем, утім навіть якщо це і стане можливим, то вкрай важким виявиться процес комунікації самих користувачів щодо домовленості використання термінів. Необхідні чітко визначені протоколи взаємодії різних укладачів онтологій. Запровадження ідеї груп за інтересами також проблемне, адже інші користувачі, маючи іншу модель світу, не в змозі розпізнати потрібні їм документи. Взагалі, розробка онтологій потребує доволі багато часу [283; 113; 239].

Урешті-решт, можна констатувати, що залучення систем, заснованих на знаннях, компонентом яких є онтології, призведе до використання інформації на якісно новому рівні.

Резюмуючи вищевикладене, зазначимо, що онтологія може мати різні форми, і для її побудови потрібне формальне декларативне подання чітко організованих конструкцій, які міститимуть словник термінів тематичної галузі, опис визначень цих термінів, класифікації чинних, теоретично можливих і неможливих взаємозв'язків між ними (у вигляді накладених обмежень).

Для опису онтологій залучають найрізноманітніші формальні мови [230; 233, ст. 199–220; 234; 278, ст. 1–25]. Усі вони можуть бути поділені на дві групи. Перша – традиційні мови опису онтологій (Interlinguas, Сус); мови, засновані на дескриптивних логіках (LOOM); мови, засновані на фреймах (OKBC, OCML, Flogic). Друга група містить мови, які ґрунтуються на Web-стандартах (XOL, UPML, SHOE, RDF з RDFS, DAML, OIL, OWL) і створені спеціально для використання онтологій в Інтернеті. Всі вони різняться потенційними можливостями опису ПГ та певною специфікою механізму логічного висновку [41; 66; 78].

Існують різноманітні методології формальної специфікації систем. З кінця 1980-х років для специфікації і проектування більших автоматизованих систем долучено методологію IDEF (Integration Definition Metodology – Об'єднання методологічних понять). Вона охоплює низку методів для моделювання і вигідно відрізняється від інших аналогів – простота в засвоєнні; надання графічних оглядових специфікацій; здатність до автоматизації; відображення різних взаємозв'язаних і необхідних для проектування аспектів побудови систем; незалежність уточнювань специфікації від проблемної сфери систем [149, ст. 194].

У цій специфікації нас цікавить метод IDEF5 – (Ontology Description Capture), який дозволяє розробляти, вивчати й підтримувати онтологію системи, що моделюється, створюючи при цьому коректні ситуації у сфері знань і погоджені висновки.

Під онтологією тут розуміють каталог термінів галузі знань, правила комбінування термінів, що відповідає прийнятому в нашому дослідженні визначенню онтології. Для підтримки процесу побудови онтологій в IDEF5 існують спеціальні онтологічні мови: схематична мова (Schematic Language-SL) й мова доробок і уточнень (Elaboration Language-EL). SL є наочною графічною мовою, призначеною для викладу фахівцями в певній галузі системи основних даних у формі онтологічної інформації (різноманітні типи діаграм і схем). Ця нескладна мова дозволяє в природний спосіб представляти

основну інформацію в початковому розвитку онтології і доповнювати чинні онтології новими даними. EL являє собою структуровану текстову мову, яка детально формулює елементи онтології.

Водночас розвиток Інтернету останнім часом зорієнтовано на розробку семантичної мережі (Semantic Web), яка залучає мову XML (Extensible Markup Language) для визначення власної структури документів і мову RDF (Resource Definition Framework), яка створює зручне середовище формалізації метаданих і відомостей щодо контексту. RDF – феномен консорціуму W3C – призначено для опису метаданих, є підмножиною мови XML, має власну мову RDF Schema для опису структури документів. Утім RDF чи не найбільш низькорівневий з наявних мов опис метаданих, оскільки оперує лише поняттями зв'язків примітивних сутностей, приміром, «об'єкт А володіє суб'єктом Б». Згодом розробники Семантичної мережі упевнилися, що засобів XML і RDF для подання інформації і метаданих для побудови повноцінної семантично зв'язаної мережі бракує. Тоді консорціумом W3C було започатковано мову онтологій OWL (Web Ontology Language), покликану, за задумом фахівців, запустити автоматизовані інструменти для глобальної мережі нового покоління [125], що є важливим аргументом на користь використання цієї мови при розробленні сучасної системи реферування та подальшого її застосування.

Мова OWL записує онтологію, зрозумілою для користувача мовою, у форматі XML. Правила онтології нею можна записати приблизно так: «Елемент А входить у колекцію Р, а колекція Р є власністю об'єкта G». Завдання OWL – уможливити інтелектуальним агентам, знайомим з конструкціями онтології, роботу з документами. Такі агенти покликані до безпосереднього контакту з користувачем і містять системи логічного висновку, які оперують конструкціями типу «Петя і Маша мешкають у будинку № 7». Звідси правила, задані в OWL, придатні лише для спеціальних систем висновку, знайомих з мовою опису семантики, і здатних обчислювати логічні висновки, на зразок мови Пролог.

Крім того, мова OWL добре підходить для опису різних нетекстових об'єктів, наприклад, колекцій зображень, відео- або аудіофайлів. З нею можна спроектувати систему опису змісту мультимедіафайлів і потім здійснювати пошук серед відеофайлів за змістом, що також є важливим для подальших розробок у галузі реферування, оскільки дозволить опрацьовувати різноманітну інформацію.

Пам'ятаємо, що документація буває різних типів, кожний з яких має свою ієрархічну структуру, свій набір перехресних посилань. Отже, онтології тут доречні для побудови єдиної інформаційної моделі, щоб у будь-якій документації вживалися певні терміни.

Також сюди варто додати проблему неможливості уживання єдиної термінології. У випадку з OWL достатньо створити декілька онтологій термінів для різних типів користувачів, але з посиланнями на відповідні терміни з інших онтологій, аби одні терміни могли автоматично транслюватися в інші, що є вагомою перевагою цієї мови.

OWL надає три підмножини, які мають різний ступінь деталізації і проектування для застосовування певними співтовариствами розробників і користувачів.

1. *OWL Lite* призначений для користувачів або додатків, які потребують лише класифікаційної ієрархії і деяких простих умов погодженості обмежень. Ця мова сприяє швидкій міграції тезаурусів та інших таксономій.

2. *OWL DL* (Description Logic) розрахований на користувачів, зацікавлених у максимальному ступені виразних можливостей мови без утрати повноти обчислень і можливості розв'язання в певний час. Рівень OWL DL орієнтований на наявні сьогодні системи опису знань і системи логічного програмування.

3. *OWL Full* розрахований на користувачів, яким необхідні максимально виразні можливості мови й синтаксична воля вибору кінцевого формату RDF без будь-яких гарантій обчислювальної повноти й можливості розв'язання. OWL Full розширює зміст терміна, залученого з заданого словника, і додає

його в онтологію, клас може одночасно розглядатися як сукупність індивідів і як індивід.

Кожен із різновидів мови OWL – розширення його більш простого попередника щодо виразних можливостей і можливостей висновків, які надаються.

Перш ніж вдаватися до будь-якої множини термінів для формального подання мовою OWL, слід точно вказати словники. Назва або посилання на конкретний словник термінів (онтологію) записується у форматі URI (unified resource identifier). Це зручно, оскільки уникнуто побудови додаткової бази даних, яка містить поняття всіх онтологій – роль бази даних унікальних ідентифікаторів покладено на розподілену базу даних доменних понять. Варто підкреслити, що в базі не зберігається інформація щодо онтології – зручна нагода уникнути колізій понять. Це ж ім'я служитиме множиною понять XML для уживаної онтології, тому в кореневому XML-елементі файлу онтології <rdf:RDF> оголошуються всі залучені онтології як множина понять.

Після того, як визначено множину понять для онтологій, включається набір припущень щодо онтології, яка описується, згрупованої всередині елемента <owl:Ontology>. Туди ж можна включити теги з коментарями, номерами версій та імпортом інших онтологій. Елемент <owl:Ontology> убирає в себе всю метаінформацію про документ.

Відмінність онтології від схеми XML полягає в тому, що це подання знання, а не формат повідомлень, як це відбувається в більшості Web-стандартів (комбінація форматів повідомлень і специфікація протоколів).

У результаті проведеного аналізу ми зупинили свій вибір на мові опису онтологій OWL, яку було застосовано для описування онтології верхнього рівня (категорії реферативних конструкцій) у розробленій нами системі автоматичного реферування.

3.3. Створення онтологій для системи автоматичного реферування

Перш ніж перейти до розроблення онтологій, слід описати, що ми розуміємо під моделлю онтології.

Стосовно визначення онтології у роботі [41], узятого за основу в нашому дослідженні, поняття онтології охоплює визначення і вживання трьох взаємозалежних компонентів: таксономії термінів, визначень термінів і правил їхнього опрацювання. У зв'язку з цим під моделлю онтології розуміємо впорядковану трійку виду $Ont = \langle X, R, F \rangle$, де X – скінченна не порожня множина концептів (понять, термінів) предметної галузі, яку представляє онтологія, R – скінченна множина відношень між концептами, F – скінченна множина функцій інтерпретації, заданих на концептах і відношеннях онтології Ont .

Як видно з визначення множин X, R, F , всі вони є скінченними і тільки X – завжди не є порожня множина. Множини R і F не завжди не порожні. У випадку, коли R і F порожні, онтологія Ont трансформується у вироджену онтологію і являє собою простий словник V : $Ont = V = \langle X, \{\}, \{\} \rangle$. Така онтологія може бути корисна для специфікації, поповнення й підтримки словників ПГ. До речі, добре відомими прикладами онтологій цього типу є індекси машин пошуку інформації в мережі Інтернет.

У запропонованій нами моделі онтологія-словник використовується для опису категорій реферативних конструкцій і є набором понять нижнього рівня з ієрархії понять верхнього рівня (підрозділ 3. 2), які мають у системі реферування однозначну інтерпретацію. Модель реферування передбачає також наявність онтологій ПГ: $Ont = T^o = \langle X, \{is_a\}, \{\} \rangle$. Відношення «is_a» має фіксовану заздалегідь семантику і дозволяє організувати структуру понять онтології у вигляді дерева. Такий підхід має свої переваги й недоліки, втім загалом є адекватним і зручним для представлення ієрархії понять.

Модель онтології надає можливість:

- представляти множину концептів X у вигляді мережевої структури;

– залучати множину відношень R , які охоплюють не тільки таксономічні відношення, а й відношення, що відбивають специфіку конкретної ПГ, а також засіб розширення множини R ;

– використовувати декларативні й процедурні інтерпретації та відношення, включно з можливістю визначення нових інтерпретацій.

Як зазначалося раніше, для вирішення задачі подання знань у системі автоматичного реферування виникла необхідність створення декількох онтологій, які сукупно дають можливість змістового конструювання реферату (через більш точне виділення в текстовій базі необхідних для цього іменникових груп).

Онтологія верхнього рівня перебуває над онтологіями ПГ і є самостійною, не залежить від ПГ і конкретної задачі, оскільки описує загальні поняття (простір, час, матерія, об'єкт, подія, дія, результат тощо; в нашому випадку – об'єкт, результат, мета й засіб).

У запропонованій нами моделі такий словник використовується для опису категорій реферативних конструкцій. Він є виродженою онтологією і являє собою кінцеву множину понять верхнього рівня, що відображають змістову структуру рефератів.

Проміжна онтологія загальнонаукової лексики, що містить загальні поняття й відношення для різних ПГ. Певною мірою вона запроваджується як посередня ланка між онтологіями ПГ.

Онтології предметних галузей містять поняття певної сфери знань або сфер, які входять до неї, і складаються з об'єктів та зв'язків між ними, описаних у термінології конкретної ПГ.

Онтології-словники не вводять експліцитного значення термінів, але в тих випадках, коли терміни належать до досить вузького словника ПГ (технічного, лінгвістичного, медичного, економічного тощо), тобто їхні значення чітко визначені, такі онтології можуть використовуватися на практиці в ролі опису ПГ.

Зв'язок категорій з онтології верхнього рівня з об'єктами з онтології ПГ описується формалізмами, які задають принципи віднесення до цих категорій об'єктів світу.

Етап концептуалізації включає витяг термінів з текстів і розподіл їх за категоріями, на базі чого і будується концептуальна модель ПГ. При цьому ми розглядаємо не всі категорії, що входять до онтології верхнього рівня, а лише ті, які зустрічаються в рефератах. У результаті в процесі побудови моделі реферату ми одержали вироджену онтологію у вигляді словника змістових категорій, притаманних рефератам – *об'єкт, результат, мета, засіб*.

При побудові онтології ПГ ми відштовхувалися від того, що в реальному світі є об'єкти, яким притаманні властивості й значення [84, ст. 20–27]. А побудова онтології ПГ зводиться до концептуалізації світу в поняттях словника для об'єктів, їхніх якісних характеристик, відмінних рис тощо. Поняття, визначені в такий спосіб у словнику, становлять прийняту в ПГ термінологію.

Як правило, онтологія містить таксономію понять, що успадковують властивості одного (проста таксономія) або ряду більш загальних понять (складна таксономія). Поняття пов'язані відношеннями з основними в системі поняттями або з невід'ємними типами. Відношення зазвичай мають кілька персональних властивостей, таких як ім'я відношення і його опис, закріплений документально. Для системи відношень установлюються семантичні обмеження.

Іншими словами, онтологічне (або концептуальне) моделювання має справу з питанням: як декларативним чином, який припускає повторне використання, описати ПГ і як обмежити використання цих даних для вирішення конкретного завдання аналізу.

Для розв'язання цього завдання модель ПГ у нашому дослідженні базується на поняттях верхнього рівня (*об'єкт, результат, мета й засіб*), які потім співвідносяться з поняттями з конкретної ПГ. Онтологія такого типу є простою таксономією.

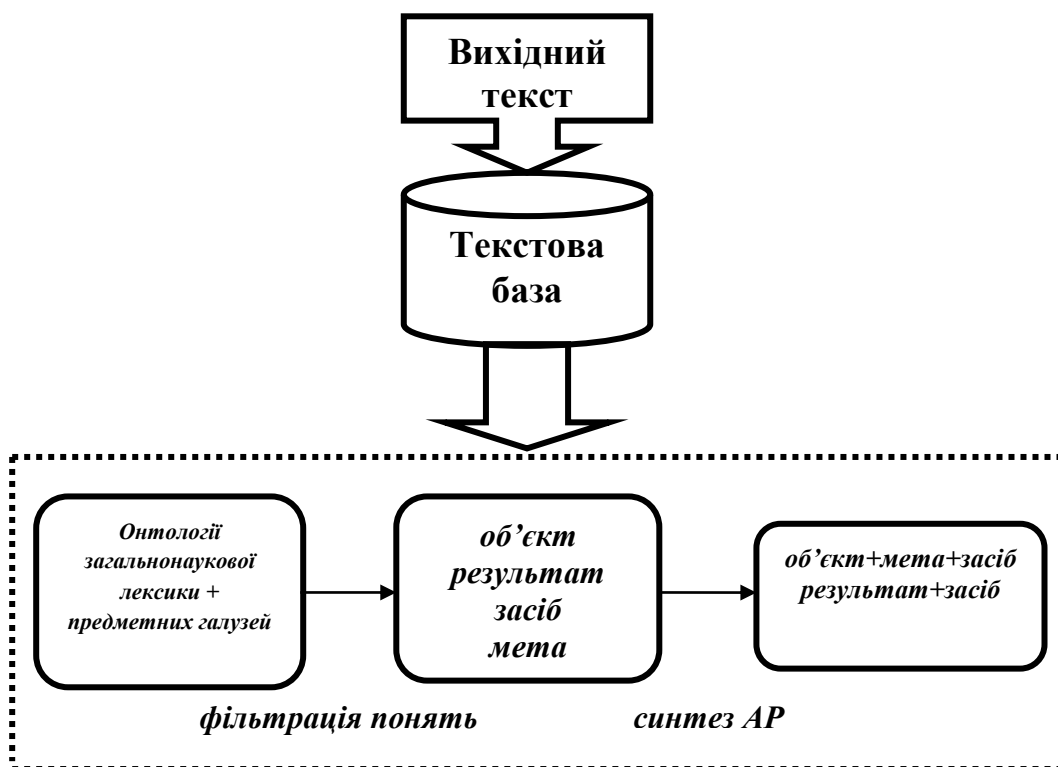
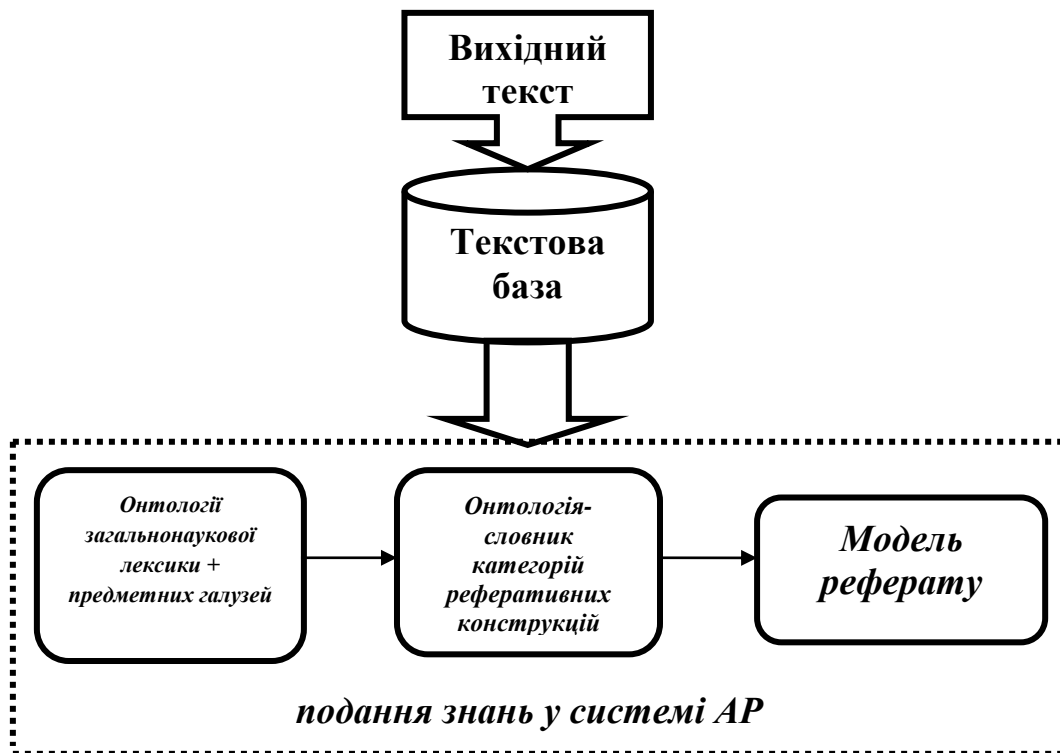


Рис. 3. 4 Подання знань у системі автоматичного реферування

У кожній прикладній онтології поняття ПГ мають яскраво виражену специфіку. Вихідна множина понять формується в результаті відбору характерних предметів і явищ, після чого будуються дерева класифікації

концептів. При цьому під терміном розуміють слово (або сполучення слів), що є точним позначенням будь-якого поняття з галузі науки й техніки, а поняття інтерпретується як логічно оформлена думка про клас предметів і явищ. Тобто в понятті відбивається сукупність істотних ознак, кожна з яких, узята окремо, необхідна, а всі разом – достатні, аби з їхньою допомогою можна було виділити даний предмет серед усіх інших і узагальнити однорідні предмети в клас. Предмети і явища, об'єднані в один клас, становлять обсяг відповідного їм поняття.

При класифікації предмети і явища, які входять до (обсягу) поняття верхнього рівня, діляться на підкласи, що утворюють нові поняття, котрі є частиною поняття верхнього рівня, і так далі до рівня конкретних термінів ПГ.

У цьому аспекті таксономія розглядається як синонім систематики – надійної і зручної системи реєстрації розмаїтості предметів і явищ шляхом розподілу за класами на підставі їхньої подібності й спорідненості в найширшому сенсі.

На рис. 3. 5 запропоновано в рамках даного підходу процедуру фільтрації термінів з текстової бази знань на етапі змістового конструювання реферату. В основу процедури покладені попередньо описані структури знань, що вказують системі реферування, які поняття є характерними для рефератів і які концептуальні властивості має те чи інше поняття.

Ці структури є виродженою онтологією у вигляді словника категорій реферативних конструкцій та онтології ПГ у вигляді таксономії понять конкретної галузі знань.

У результаті зіставлення термінів із текстової бази знань з цими онтологіями формується набір понять, необхідний для змістового конструювання реферату. Для цього виділені на етапі аналізу заголовка і вихідного тексту терміни T_{1m} , T_{2m} , T_{3m} з текстової бази зіставляються з термінами T_1 , T_2 , ..., T_N з онтології заданої ПГ. Якщо відповідність не встановлена (T_{2m}), термін виводиться з подальшого розгляду. Слова, що залишилися (T_1 , T_5), перевіряються на наявність у словнику категорій

реферативних конструкцій i – в разі виконання цієї умови ($T_i - \Pi_{2i}$) – термін, що аналізується (T_i), підставляється в модель реферату відповідно до правил заповнення актантних структур реферативних речень, описаних у моделі реферату.

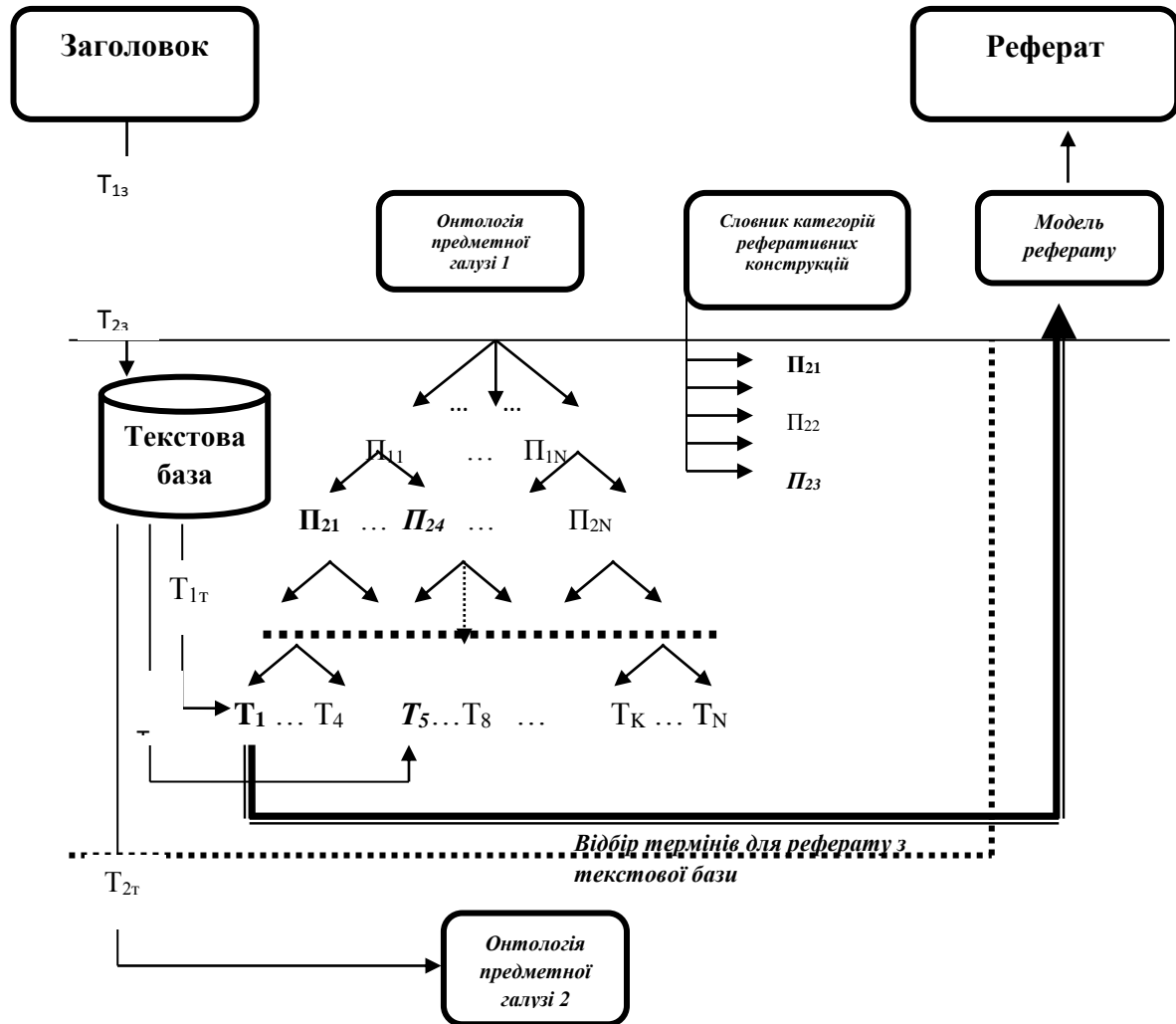


Рис. 3. 5 Схема реферування на основі онтологій

Для практичної реалізації розробленої схеми реферування необхідно формалізувати опис роботи системи мовою, прийнятною для машини. Як було позначено в підрозділі 3. 2, при виборі способу формалізації ми орієнтувалися на технології Semantic Web, а саме на мову опису словників RDF-термінів (RDFS – RDF Schema), який надає можливість машинної інтерпретації даних і об’єднання інформації.

RDF-схеми дозволяють визначати семантику даних, слугують фундаментом для більш багатих мов опису моделей ПГ (мов опису онтологій),

які дозволяють адаптувати системи до Інтернет і семантично обробляти дані. Мета таких мов – показати додаткову семантику ресурсів, яка інтерпретується машиною, тобто зробити машинне подання даних більш схожим на стан речей у реальному світі.

За допомогою мови OWL, ми описали онтологію верхнього рівня та її зв'язок з онтологіями ПГ.

Онтологія верхнього рівня в нашій системі є набором категорій: об'єкт, результат, мета й засіб, які описуються в такий спосіб:

```
<owl:Class rdf:ID=«Об'єкт»/>
```

```
<owl:Class rdf:ID=«Результат»/>
```

```
<owl:Class rdf:ID=«Мета»/>
```

```
<owl:Class rdf:ID=«Засіб»/>
```

Синтаксис `rdf:ID=«Об'єкт»` залучається, щоб увести назву, як частину його визначення. За допомогою «#об'єкт» можна посилатися на клас «Об'єкт» усередині документа, наприклад, `rdf:resource=«#об'єкт»` із інших онтологій, уживаючи повну форму: `http://www.XXX/екземпляр#об'єкт` або `http://www.XXX/послідовність#об'єкт`.

Визначення можуть розширюватися (в нашому прикладі, аби більш конкретизувати поняття «Об'єкт») і розподілятися (аби конструювати більші онтології з блоків).

З цією метою застосовується синтаксис `rdf:about=«#об'єкт»`. Залучення синтаксису `rdf:about=«&ont;#x»` – ключовий елемент у створенні розподіленої онтології. Цей опис дозволяє розширити імпортоване визначення *x*, не змінюючи оригінал документа, і конструювати нові онтології з наявних блоків.

Для використання описаного класу іншим документам знадобиться посилання на даний клас. Це забезпечується описом `NAMESPACE` і визначенням `ENTITY`, які містять документ, що визначається, як джерело:

```
<!ENTITY sem «http://www.XXX/семантика#» >
```

```
<rdf:RDF xmlns:sem =«http://www.XXX /семантика#» >
```

Уклавши ці визначення, ми можемо звертатися до класу «Об'єкт», використовуючи або XML-тег «sem: Об'єкт», або значення атрибута «&sem; Об'єкт». Крім того, завжди можна зіслатися на ресурс, використовуючи його повний URI, у нашому випадку «http://www.XXX/семантика#об'єкт».

Фундаментальним таксономічним конструктором для класів є `rdfs:subClassOf`. Він зв'язує більш частковий клас із більш загальним класом. Якщо X – підклас Y , то кожний представник X – також представник Y . Відношення «`rdfs:subClassOf`» є транзитивним. Якщо X – підклас Y , і Y – підклас Z , то X – підклас Z .

```
<owl:Class rdf:ID=«Семантика»>
  <rdfs:subClassOf rdf:resource=«# об'єкт»/>
  ...
</owl:Class>
```

Ми визначаємо «Семантику» як підклас класу «Об'єкт» із онтології верхнього рівня.

Зазначимо, що переважна більшість зусиль із розробки онтології присвячена зв'язуванню класів і властивостей таким чином, аби максимально точно передати закладений у поняття зміст. Це найскладніша частина розроблені онтології. У зв'язку з цим привабливим видається злиття воедино вже наявних онтологій. У цьому разі допомога спеціального описаного інструментарію буде необхідна для підтримки узгодженості.

Наведене визначення семантики, звичайно, ще дуже неповне. Далі нам необхідно виділити екземпляри цього класу й описати судження про них. Для визначення екземпляра досить оголосити його членом якогось класу.

```
<Семантика rdf:ID= «Смысл»/>,
що є ідентичним за значенням запису
<owl:Thing rdf:ID=«Смысл»/>
<owl:Thing rdf:about=«#Смысл»>
  <rdf:type rdf:resource=«#семантика»/>
</owl:Thing>
```

rdf:type – це властивість RDF, що зв’язує екземпляр із класом.

На цьому принципі побудований опис онтологій у розроблювальній нами моделі автоматичного реферування.

Побудовані онтології є лише першим кроком до створення відповідної системи подання знань у системі автоматичного реферування. На нашу думку, кращі результати можливі лише за наявності повних онтологій термінологічної і загальнонаукової лексики, а також за наявності в системі АР текстової бази у вигляді «відносно простої і незмінної семантичної конфігурації», яка дозволяє досить швидко знаходити змістові аспекти для заповнення моделі реферату преференційними значеннями мовних одиниць тексту, які характерні для наукової і технічної літератури [60]. Створення такої текстової бази передбачає дослідження процесів розуміння тексту, заснованих на лінгвістичному аналізі структури зв’язного тексту. Подальше дослідження в цьому напрямку дозволить побудувати повноцінну систему автоматичного реферування з опертям на знання.

Висновки

Під час дослідження й аналізу емпіричних даних ми дійшли ряду висновків і узагальнень, як-от:

1. Запропоновано підхід до опису семантичних зв'язків «Текст-Реферат» у вигляді семантико-контекстної моделі реферування, що забезпечує розуміння тексту і його трансформацію зі збереженням смислу за рахунок використання:

– заголовка, який виражає зміст вихідного тексту в концентрованому вигляді;

– текстової бази, яка містить речення, що являють собою «інформаційне ядро» тексту й відображають його тематику;

– онтологій, що містять незалежну від тематики тексту інформацію: онтологій предметних галузей, онтології загальнонаукової лексики й онтології верхнього рівня.

2. Аналіз змісту тексту під час реферування передбачає аналіз заголовка, тому що ми розглядаємо останній як реферат мінімального обсягу або як текст із максимальним рівнем згортання змісту.

3. У результаті дослідження змістової і синтаксичної структури заголовка виявилася його схожість зі структурою реферату. За результатами порівняльного аналізу з текстом реферату встановлено, що всі компоненти структури заголовка відповідають компонентам структури реферату.

4. Проведені дослідження дозволили простежити чітке продовження процедури згортання (компресії) в заголовку перед процедурою компресії в рефераті як на семантичному, так і на синтаксичному рівнях.

5. При збереженні тих самих змістових складових у рефераті й у заголовку вони мають при цьому синтаксичні й граматичні особливості їх вираження, які підвищують рівень компресії в заголовку.

6. Було побудовано класифікацію лексем заголовка, з урахуванням якої на підставі проведеного дослідження змістової і синтаксичної структури заголовка було побудовано загальну модель заголовку.

7. Обґрунтована необхідність наявності в системі реферування текстової бази, до якої входять речення, що містять слова з заголовка або їх смислові еквіваленти з тексту і дозволяють здійснити перехід від заголовка до тексту, а потім до реферату під час змістового аналізу тексту й семантичного наповнення моделі реферату.

8. Запропоновано алгоритм побудови текстової бази знань на цьому етапі досліджень на основі понять, які містяться в заголовку документа, і слів-указівників у тексті на змістові аспекти, необхідних для побудови реферату.

9. Запропоновані для реферування модель компресії тексту й модель реферату, за своєю суттю, еквівалентні до застосування макроправил для побудови читачем макроструктур як однієї зі стратегій розуміння дискурсу, що дозволяє побудувати формальний перехід від вихідного тексту до реферату, який складається з кількох речень.

10. Для оптимізації процедури вибору речення з тексту в текстову базу обґрунтовано необхідність наявності в системі автоматичного реферування онтологій, в яких достеменно зафіксовані всі концепти відповідної предметної галузі.

11. Розглянуто онтології як один з найбільш перспективних способів подання й опрацювання знань і запропоновано здійснювати подання знань у системах АР шляхом створення онтологій.

12. Проведено аналіз мов опису онтологій. Обрано мову OWL для описання онтологій у системі автоматичного реферування.

13. Для розв'язання завдань презентації знань у системі автоматичного реферування обґрунтовано необхідність створення декількох онтологій, які сукупно дають можливість змістового конструювання реферату.

14. Побудовано онтологію верхнього рівня, яка є виродженою онтологією, у вигляді словника метазначень (змістових категорій, притаманних рефератам, – об'єкт, результат, мета, засіб).

15. Підготовлено вироджену онтологію загальнонаукової лексики для побудови проміжної онтології загальнонаукової лексики, яка дозволить

зв'язати загальнонаукові поняття з поняттями різноманітних предметних галузей.

16. Побудовані онтології та запропонований алгоритм побудови текстової бази є лише першим кроком до створення відповідної системи подання знань у системі автоматичного реферування.

ПІДСУМОК

У результаті проведеного дослідження запропоновано підхід до розробки системи автоматичного реферування з опорою на знання, згідно з яким моделювання процесу реферування, яке є сукупністю складних процесів розуміння й компресії змісту. Досліджено результат цих процесів – реферат, причому не розгорнутий, інформативний, а стислий, індикативний, як об'єкт найбільш простий за формою, але такий, що відображає основні змістові особливості реферативного тексту.

У межах запропонованого підходу здійснено дослідження загальних закономірностей для всіх видів стиснення тексту, тобто свідомо припущено теоретичну та емпіричну неповноту на початковому етапі, що дозволило, відштовхнувшись від розуміння вивчених на початковому етапі роботи закономірностей, надалі розширити сферу дослідження і звести моделювання процесу реферування до кількох самостійних, однак взаємозумовлених завдань, розв'язання яких здійснювалося поетапно.

На першому етапі побудовано уніфікований реферат у вигляді набору речень із семантикою об'єкта, результату, мети й засобу та розроблено методику формування ознак, які характеризують реферат як результат компресії (і узагальнення) інформації на синтаксичному й семантичному рівнях. На базі проведеного дослідження синтаксичної структури реферативних речень і речень первинних текстів побудовано модель компресії тексту на синтаксичному рівні. А проведений аналіз механізмів стиснення тексту на семантико-синтаксичному, власне семантичному й лексико-семантичному рівнях дозволив побудувати модель компресії тексту на семантичному рівні дозволив побудувати модель реферату у вигляді типових для індикативних рефератів семантико-синтаксичних структур.

Крім того зроблено класифікацію лексики, використовуваної в індикативних рефератах і описані правила наповнення синтаксичних конструкцій реферату поняттями, що узагальнюють семантику вихідного тексту.

Все це дозволило розробити першу версію комп'ютерної програми «АвтоРеферат», яка підтвердила правильність породження реферативних речень відповідно до розробленої моделі реферування й водночас указала на очікувану змістову неповноту цих речень і необхідність більш глибокого змістового аналізу первинного тексту.

З цією метою на другому етапі досліджень було вивчено семантичні відношення вихідного тексту й реферату і запропоновано підхід до опису семантичних зв'язків «Текст-Реферат» у вигляді семантико-контекстної моделі реферування, яка забезпечує розуміння тексту за рахунок використання моделі заголовка, текстової бази та онтологій.

Для побудови семантико-контекстної моделі реферування розглянуто заголовки як реферат мінімального обсягу або як текст із максимальним рівнем згортання змісту, на підставі чого проведено дослідження змістової і синтаксичної структури заголовка й виявлено його схожість зі структурою реферату. Проведено класифікацію лексем заголовка, з урахуванням якої побудовано загальну модель заголовка.

Крім того, обґрунтовано необхідність наявності в системі реферування текстової бази, до якої входять речення, що містять слова із заголовка або їх смислові еквіваленти з тексту й слова-указівники на відповідні змістові аспекти в тексті. Запропоновано алгоритм побудови текстової бази з використанням заголовка й слів-указівників, що є необхідним для створення реферату.

Для оптимізації процедури вибору речень із тексту в текстову базу обґрунтовано необхідність наявності в системі автоматичного реферування декількох онтологій: онтології верхнього рівня, онтології загальнонаукової лексики та онтологій предметних галузей. Для опису онтологій обрано мову OWL з метою забезпечення можливості застосування розробленої системи реферування в мережі Інтернет.

Теоретичне значення виконаного дослідження полягає в тому, що побудова моделі семантичних зв'язків «Текст-Реферат» з використанням

заголовка й текстової бази, яка являє собою основу макроструктури тексту, дозволило зробити висновки про те, як відбувається змістове згортання у процесі реферування, і наблизитися до формального опису певних аспектів процесу розуміння тексту людиною.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Автоматизация анализа научного текста. – К. : Наук. думка, 1984. – 257 с.
2. Андреев Н. Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении / Н. Д. Андреев. – Л. : Наука, 1967. – 404 с.
3. Анисимов А. В. Компьютерная лингвистика для всех: Мифы. Алгоритм. Язык / А. В. Анисимов. – К. : Наук. думка, 1991. – 208 с.
4. Апресян Ю. Д. Идеи и методы современной структурной лингвистики (краткий очерк) / Ю. Д. Апресян. – М. : Просвещение, 1966. – 304 с.
5. Апресян Ю. Д. Лингвистическое обеспечение системы ЭТАП-2 / Ю. Д. Апресян, И. М. Богуславский, Л. Л. Иомдин и др. – М. : Наука, 1989. – 296 с.
6. Бездушный А. А. RDFS как основа среды разработки цифровых библиотек и Web-порталов / А. А. Бездушный, А. Н. Бездушный, А. К. Нестеренко и др. // Электронные библиотеки. – 2003. – Том 6. – Вып. 3. – 168 с.
7. Белоногов Г. Г. Автоматизация процессов накопления, поиска и обобщения информации / Г. Г. Белоногов, А. П. Новоселов. – М. : Наука, Главная редакция физико-математической литературы, 1979. – 256 с.
8. Белоногов Г. Г. Автоматизированная обработка научно-технической информации. Лингвистические аспекты / Г. Г. Белоногов, Б. А. Кузнецов, А. П. Новоселов. – М. : ВИНТИ, 1984. – 316 с.
9. Белоногов Г. Г. Автоматизированные информационно-поисковые системы / Г. Г. Белоногов, Р. Г. Котов. – М. : Сов. радио, 1968. – 184 с.
10. Белоногов Г. Г. Автоматизированные информационные системы / Г. Г. Белоногов, В. И. Богатырев. – М. : Сов. радио, 1973. – 328 с.

11. Белоногов Г. Г. Языковые средства автоматизированных информационных систем / Г. Г. Белоногов, Б. А. Кузнецов. – М. : Наука, 1983. – 288 с.
12. Бельская И. К. Язык человека и машина / И. К. Бельская. – М. : Изд-во Московского ун-та, 1969. – 410 с.
13. Березин Ф. М. Лингвистическое обеспечение информационных систем / Ф. М. Березин // Сборник научно-аналитических обзоров. Серия: Теория и история языкознания. – М. : ПИК ВИНТИ, 1987. – 220 с.
14. Берзон В. Е. Синтаксические сверхфразовые связи и их инженерно-лингвистическое моделирование / В. Е. Берзон, А. Б. Брайловский. – Кишинев: Штиинца, 1984. – 167 с.
15. Берзон В. Е. Классификация коннекторов и диалоговые системы автоматического реферирования / В. Е. Берзон, А. Б. Брайловский // НТИ. Сер. II. – 1979. – № 11. – С. 19–23.
16. Бессонов А. В. Фрегевская концептуализация логико-семантической теории / А. В. Бессонов // Концептуализация и смысл. – Новосибирск, 1990. – С. 20–30.
17. Бир Ст. Кибернетика и управление производством: Пер. с англ. / Ст. Бир. – М. : Наука, 1965. – 392 с.
18. Биро Ю. Исследование и разработка методов автоматического реферирования научных статей по экономике: Дис. канд. пед. наук: 05.25.03 / Ю. Биро. – Л. : Наука, 1990. – 280 с.
19. Бирюков Б. В. Кибернетика в гуманитарных науках / Б. В. Бирюков, Е. С. Геллер. – М. : Наука, 1973. – 383 с.
20. Блохин Ю. И. Классификация и кодирование технико-экономической информации / Ю. И. Блохин. – М. : Экономика, 1976. – 192 с.
21. Блюменау Д. И. Проблемы аннотирования и реферирования научной литературы / Д. И. Блюменау, В. П. Леонов, Н. Г. Чагина // Труды ЛГИК им. Н. К. Крупской. – Л. : Наука, 1982. – 125 с.

22. Блюменау Д. И. Проблемы свертывания научной информации / Д. И. Блюменау. – Л. : Наука, 1982. – 166 с.
23. Блюменау Д. И. Формализованное реферирование с использованием словесных клише (маркеров) / Д. И. Блюменау, Н. И. Гендина, И. С. Добронравов и др. // НТИ. – Сер. 2. – 1981. – № 2. – С. 16–20.
24. Блюменау Д. И. Функции обзоров в системе научных коммуникаций / Д. И. Блюменау // Труды ЛГИК им. Н. К. Крупской. – Л. : Наука, 1976. – Т. 29. – С. 233–244.
25. Богданов В. В. Прикладная лингвистика и теория языка / В. В. Богданов, Л. В. Бондарко, В. Д. Буторов и др. // Структурная и прикладная лингвистика. – Л. : ЛГУ, 1987. – С. 3–16.
26. Бойко Е. И. Механизмы умственной деятельности (Динамические временные связи) / Е. И. Бойко. – М. : Педагогика, 1976. – 248 с.
27. Брудный А. А. Понимание и текст / А. А. Брудный // Загадка человеческого понимания. – М. : Политиздат, 1991. – С. 114–128.
28. Брунер Дж. Психология познания. За пределами непосредственной информации: Пер. с англ. / Дж. Брунер. – М. : Прогресс, 1977. – 413 с.
29. Бусленко Н. П. Моделирование сложных систем / Н. П. Бусленко. – М. : Наука, 1978. – 400 с.
30. Буторов В. Д. Приближенные методы анализа языка в прикладной лингвистике / В. Д. Буторов // Структурная и прикладная лингвистика. – Л. : ЛГУ, 1983. – Вып. 2. – С. 131–138.
31. Быстрицкий Е. К. Понимание – окно в новый мир? / Е. К. Быстрицкий // Загадка человеческого понимания. – М. : Политиздат, 1991. – С. 221–234.
32. Васильев Л. Г. Текст и его понимание / Л. Г. Васильев. – Тверь: ТГУ, 1991. – 678 с.
33. Вейзе А. А. Реферирование текста / А. А. Вейзе. – Мн. : Изд-во БГУ, 1978. – 128 с.

34. Вейценбаум Дж. Возможности вычислительных машин и человеческий разум. От суждений к вычислениям: Пер. с англ. / Дж. Вейценбаум. – М. : Радио и связь, 1982. – 368 с.
35. Викери Б. Фасетная классификация: Руководство по составлению и использованию отраслевых схем / Б. Викери. – М. : ГПНТБ, 1973. – 73 с.
36. Виноград Т. Программа, понимающая естественный язык / Т. Виноград. – М. : Мир, 1976. – 295с.
37. Витгенштейн Л. Логико-философский трактат / Л. Витгенштейн. – М.: Изд. иностр. лит., 1958. – 133 с.
38. Вопросы информационной теории и практики. – М.: ВИНТИ. – 1974. – № 25. – С. 66–67
39. Воробьев Г. Г. Документ: информационный анализ / Г. Г. Воробьев. – М. : Мир, 1973. – 220 с.
40. Гаврилов Л. А. Реферирование как специализированная речевая деятельность / Л. А. Гаврилов // Система языка и перевод. – М. : Изд-во МГУ, 1983. – С. 17–28.
41. Гаврилова Т. А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб. : Питер, 2000. – 384 с.
42. Гадамер Х.-Г. Истина и метод / Х.-Г. Гадамер. – М. : Наука, 1988. – 699 с.
43. Гайсарян С. С. Объектно-ориентированные технологии проектирования прикладных программных систем [электрон. ресурс] / С. С. Гайсарян. – Режим доступа: <http://sure.org.ru/docs/c/oop>.
44. Гальперин И. Р. Текст как объект лингвистического исследования / И. Р. Гальперин. – М. : Наука, 1981. – 140 с.
45. Гендина Н. И. Лингвистическое обеспечение автоматизированных библиотечных систем / Н. И. Гендина. – Алма-Ата: Гылым, 1991. – 221 с.
46. Гиндин С. И. Методы автоматического фрагментирования текста, опирающиеся на характеристики внутреннего состава фрагментов /

С. И. Гиндин // Семиотика и информатика. – М. : ВИНТИ, 1977. – Вып. 9. – С. 35–82.

47. Гиндин С. И. Проблемы анализа и синтеза целого текста в системах машинного перевода, диалоговых и информационных системах / С. И. Гиндин, Н. Н. Леонтьева. – М. : ВЦП, 1978. – 84 с.

48. Глушков В. М. Введение в кибернетику / В. М. Глушков. – К. : АН УССР, 1964. – 324 с.

49. Гончаренко В. В. Автоматическая переработка текста (Системный подход к решению проблемы машинного перевода) / В. В. Гончаренко, Э. М. Добрускина, Н. М. Котельникова и др. – Кишинев: Штиинца, 1978. – 94 с.

50. Гончаренко В. В. Фреймы для распознавания смысла текста / В. В. Гончаренко, Е. А. Шингарева. – Кишинев: Штиинца, 1984. – 198 с.

51. Горский Д. П. Вопросы абстракции и образование понятий / Д. П. Горский. – М. : Изд-во Акад. наук СССР, 1961. – 148 с.

52. Горский Д. П. Краткий словарь по логике / Д. П. Горский. – М. : Просвещение, 1991. – 208 с.

53. Горский Д. П. Обобщение и познание. – М. : Мысль, 1985. – 152 с.

54. Горькова В. И. Пособие по реферированию и редактированию научно-технической литературы (Справочник для редакторов, референтов и корректоров) / В. И. Горькова. – М. : Наука, 1964. – 268 с.

55. Горькова В. И. Реферат в системе научной коммуникации. Направления совершенствования лингвистических и структурных характеристик / В. И. Горькова, Э. А. Борохов // Итоги науки и техники. Сер. Информатика. – Т. 11. – М. : ВИНТИ, 1987. – 232 с.

56. Грязнухина Т. А. Анализ предложных связей в научном тексте / Т. А. Грязнухина. – К. : Наукова думка, 1985. – 148 с.

57. Грязнухина Т. А. Общие принципы построения автоматического синтаксического анализатора / Синтаксический анализ научного текста на ЭВМ. – К. : Наукова Думка, 1999. – С. 6–39.

58. Губин М. В. Эффективный алгоритм формирования контекстно-зависимых аннотаций / М. В. Губин, А. И. Меркулов // Компьютерная лингвистика и интеллектуальные технологии. – М. : Наука, 2005. – С. 116–120
59. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник / Н. П. Дарчук. – К. : Київський університет, 2008. – 351 с.
60. Дейк ван Т. А. Стратегии понимания связного текста / Т. А. ван Дейк, В. Кинч // Новое в зарубежной лингвистике. – Вып. 23: Когнитивные аспекты языка. – М. – 1988. – С. 153–211.
61. Дерешко Б. Ю. Классификация информационно-коммуникационных образовательных технологий и их использование в учебном процессе / Б. Ю. Дерешко // Телекоммуникации и информатизация образования. – М. – 2006. – № 1 (32). – С. 7–12.
62. Джиго А. А. ГОСТ 7.9-95. Реферат и аннотация. Общие требования / А. А. Джиго, С. Ю. Калинин // Стандарты по издательскому делу. – М. : Юристъ, 1998. – 132 с.
63. Добрускина Э. М. Синтаксические сверхфразовые связи и их инженерно-лингвистическое моделирование / Э. М. Добрускина, В. Е. Берзон. – Кишинев: Штиинца, 1986. – 167 с.
64. Дрейфус Х. Чего не могут вычислительные машины. Критика искусственного разума: Пер. с англ. / Х. Дрейфус. – М. : Прогресс, 1978. – 335 с.
65. Дридзе Т. М. Интерпретационные характеристики и классификация текстов (с учетом специфики интерпретационных сдвигов) / Т. М. Дридзе // Смысловое восприятие речевого сообщения (в условиях массовой коммуникации). – М. : Наука, 1976. – С. 34–44.
66. Дюк В. Data Mining: Учеб. курс / В. Дюк, А. Самойленко. – СПб. : Питер, 2001. – 368 с.
67. Ермак В. Д. Классификация?.. Типология... Идентификация!.. [электрон. ресурс] / В. Д. Ермак. Сайт газеты «Соционические чтения». – 1999. – № 10 (31). – Режим доступа: <http://ru.laser.ru/socion/dn/index.html>.

68. Жданова Г. С. О научных основах методики реферирования / Г. С. Жданова // НТИ. – Сер 1. – 1967. – № 2. – С. 12–16.
69. Жинкин Н. И. Речь как проводник информации. – М. : Наука, 1982. – 153 с.
70. Жолкова А. И. К вопросу о реферировании научно-технической периодики в реферативном журнале ВИНТИ // НТИ. – Сер. 1. – 1970. – № 5. – С. 24–25.
71. Жолкова А. И. О применении методов фасетного анализа к реферированию / А. И. Жолкова // НТИ. – Сер. 1. – 1975. – № 6. – С. 26–28.
72. Загоруйко Ю. А. Описание сложных предметных областей на основе интеграции средств представления знаний / Ю. А. Загоруйко, И. Г. Попов // Труды международного семинара Диалог'97 по компьютерной лингвистике и ее приложениям. – М. : Наука, 1997. – С. 110–115.
73. Закономерности структурной организации научно-реферативного текста. – К. : Наукова думка, 1982. – 322 с.
74. Замаруєва І. Мова – інформація – знання / І. Замаруєва // Інформація та нові технології. – К. – 1996. – № 1. – С. 2–4.
75. Ильичев Л. Ф. Философский энциклопедический словарь / Л. Ф. Ильичев, П. Н. Федосеев, С. М. Ковалев и др. – М. : Сов. Энциклопедия, 1983. – 840 с.
76. Исаков В. П. О выборе алгоритмического языка для программирования задач обработки смысловой информации / В. П. Исаков, Н. Г. Лоскутов // Статистика речи и автоматический анализ текста. – Л. : НТИ, 1971. – 162 с.
77. Калиниченко Л. А. Машины баз данных и знаний / Л. А. Калиниченко, В. М. Рывкин. – М. : Наука, 1990. – 296 с.
78. Кальченко Д. Интеллектуальные агенты семантического Web'a / Д. Кальченко // Компьютер Пресс. – 2004. – № 10. – С. 26–32.

79. Карасев С. А. Построение системы автоматического реферирования, включающей анализ и синтез текста (АНАСИН) / С. А. Карасев. – Науч. техн. информация. – 1978. – Сер. 2. – № 5. – С. 22–28.
80. Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка / Ю. Н. Караулов. – М. : Наука, 1981. – 368 с.
81. Карнап Р. Преодоление метафизики логическим анализом языка / Р. Карнап // Вестник МГУ. – 1993. – Сер. 7. – № 6. – С. 11–26.
82. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: Підручник / Є. А. Карпіловська. – Донецьк: Юго-Восток, Лтд, 2006. – 188 с.
83. Китов А. И. Программирование информационно-логических задач / А. И. Китов. – М. : Сов. радио, 1967. – 328 с.
84. Клещев А. С. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия «онтология» / А. С. Клещев, И. Л. Артемьева // НТИ. – 2001. – Сер. 2. – № 2. – С. 20–27.
85. Клименко Н. Ф. Як народжується слово: Наук.-попул. вид. / Н. Ф. Клименко. – К. : Рад. шк., 1991. – 288 с.
86. Ковтун Т. В. Типы наименований текстов в научно-популярном подстиле / Т. В. Ковтун. – Автореферат, 1990. – С. 4–5.
87. Кокорєва Л. В. Діалогові системи та представлення знань / Л. В. Кокорєва, О. Л. Перевозчикова, К. Л. Ющенко. – К. : Наукова думка, 1992. – 448 с.
88. Кондаков Н. И. Логический словарь-справочник / Н. И. Кондаков. – М. : Наука, 1975. – 717 с.
89. Конюшко А. Е. Реферативные издания. Типология и редакторская подготовка / А. Е. Конюшко. – Автореферат диссертации на соискание ученой степени кандидата филологических наук. – М. : Наука, 1992. – 24 с.
90. Коробков И. А. Палеонтологические описания: Метод. пособие / И. А. Коробков. – Л. : Недра, 1971. – 200 с.

91. Котов В. Е. Сети Петри / В. Е. Котов. – М. : Наука, 1984. – 160 с.
92. Котов Р. Г. Лингвистические аспекты автоматизированных систем управления / Р. Г. Котов. – М. : Наука, 1977. – 168 с.
93. Кочерган М. П. Загальне мовознавство: Підручник. – К. : Академія, 2003. – 464 с.
94. Криницкий Н. А. Программирование и алгоритмические языки / Н. А. Криницкий, Г. А. Миронов, Г. Д. Фролов. – М. : Наука, 1979. – 510 с.
95. Кузин Л. Т. Основы кибернетики: В 2-х т. Т. 2. Основы кибернетических моделей: Учеб. пособие для вузов / Л. Т. Кузин. – М. : Энергия, 1979. – 584 с.
96. Кулагина О. С. Исследования по машинному переводу / О. С. Кулагина. – М. : Наука, 1979. – 324 с.
97. Кулик А. Н. Реферат научного документа в единой информационной сети / А. Н. Кулик // НТИ. – Сер. 1. – 1971. – № 1. – С. 28–32.
98. Курбаков К. И. Кодирование и поиск информации в автоматическом словаре / К. И. Курбаков. – М. : Сов. радио, 1968. – 248 с.
99. Лазаренко О. В. Информационные технологии переводчика. Ч.2. Реферирование: Метод. пособие: Для студентов 5 курса фак. «Референт-переводчик» / О. В. Лазаренко. – Харьков: Изд-во НУА, 1999. – 37 с.
100. Лазаренко О. В. Классификация понятий в системе автоматического реферирования / О. В. Лазаренко // Wiek XXI. – THE 21st CENTURE.– Варшава, PWSBiA, 2002. – № 4 (6) – С.189–196.
101. Лазаренко О. В. Математические модели префиксального словообразования и их использование в системах автоматической обработки текстов на русском языке: Дис... канд. техн. наук: 05.13.01. – Харьков: Изд-во НУА, 1984. – 203 с.
102. Лазаренко О. В. Аналіз смислової структури заголовка як тексту з максимальним рівнем узагальнення / О. В. Лазаренко, Т. В. Попова // Проблеми семантики слова, речення та тексту: Збірник наукових праць. – К. : КНЛУ, 2004. – Вип. 12. – С. 143–149.

103. Лазаренко О. В. Моделювання процесу узагальнення в системі автоматичного реферування / О. В. Лазаренко, А. А. Яковенко. – Харьков: Изд-во НУА, 2007. – 136 с.
104. Ландэ Д. В. Агенты новостей в сети Интернет / Д. В. Ландэ // СНІР/ Украина. – 2001. – № 5. – С. 108–111.
105. Ландэ Д. В. Поиск знаний в Internet. Профессиональная работа: Пер. с англ. / Д. В. Ландэ. – М. : Вильямс, 2005. – 272 с.
106. Левицкий Ю. А. Основы теории синтаксиса / Ю. А. Левицкий. – М. : УРСС, 2002. – 234 с.
107. Леонов В. П. О методах автоматического реферирования (США, 1958–1974 гг.) / В. П. Леонов // НТИ. – 1975. – Сер. 2. – № 6. – С. 16–20.
108. Леонов В. П. Реферирование и аннотирование научно-технической литературы / В. П. Леонов. – Новосибирск: Наука, 1986. – 172 с.
109. Лингвистические проблемы автоматизации редакционно-издательских процессов. – К. : Наукова думка, 1986. – 230 с.
110. Лингвистический энциклопедический словарь / Гл. ред. В. Н. Ярцева. – М. : Сов. Энциклопедия, 1990. – 685 с.
111. Линдсей П. Переработка информации у человека (Введение в психологию): Пер. с англ. / П. Линдсей, Д. Норман. – М. : Мир, 1974. – 550 с.
112. Лосева Л. М. К изучению межфразовой связи (Абзац и сложное синтаксическое целое) / Л. М. Лосева // Русский язык в школе, 1976. – № 1. – С. 89–94.
113. Майкевич Н. В. От информационного пространства к пространству знаний. Онтологии в Интернет [Электрон. ресурс] / Н. В. Майкевич // Труды конференции КИИ'98, Пущино (Россия). – 1998. – С. 152–158. – Режим доступа: <http://www.uchcom.botik.ru/nut/CAI98.html>.
114. Максудова М. С. Один из подходов к автоматическому индексированию рефератов / М. С. Максудова // Вопросы кибернетики. – Ташкент: Изд-во АН УзССР, 1971. – Вып. 44. – С. 190–197.

115. Мартин Дж. Организация баз данных в вычислительных системах / Дж. Мартин. – М. : Мир, 1980. – 663 с.
116. Марчук Ю. Н. Основы компьютерной лингвистики / Ю. Н. Марчук. – М. : Народный учитель, 2000. – 226 с.
117. Маторин С. И. О моделировании интеллектуального понимания языка делового общения / С. И. Маторин // НТИ. – 1997. – Сер. 2. – № 4. – С. 9–17.
118. Мельников Г. П. Системология и языковые аспекты кибернетики / Г. П. Мельников. – М. : Сов. радио, 1978. – 368 с.
119. Мельчук И. А. Опыт теории лингвистических моделей «Смысл ↔ Текст» / И. А. Мельчук. – М. : Наука, 1974. – 315 с.
120. Мидоу Ч. Анализ информационных систем / Ч. Мидоу. – М. : Прогресс, 1977. – 400 с.
121. Милль Д. С. Система логики / Д. С. Милль. – Спб. : Вольф, 1878. – Т.2. – 507 с.
122. Минский М. Фреймы для представления знаний: Пер. с англ. / М. Минский. – М. : Энергия, 1979. – 152 с.
123. Миронова В. Ю. К вопросу о жанровых особенностях языка научно-технических реферативных журналов / В. Ю. Миронова // Науч. труды МГПИИЯ им. М. Горького. – М. – 1974. – Вып. 82. – С. 182.
124. Михайлов А. И. Основы информатики / А. И. Михайлов, А. И. Черный, Р. С. Гиляревский. – М. : Наука, 1968. – 274 с.
125. Михаленко П. Язык онтологий в Web [Электрон. ресурс] / П. Михаленко // Открытые системы. – Режим доступа: osp.ru#02/2004.
126. Мурашко А. Г. Первое знакомство с сетями Петри: Учебное пособие / А. Г. Мурашко. – К. : УМК ВО, 1988. – 71 с.
127. Налимов В. В. Вероятностная модель языка. О соотношении естественных и искусственных языков / В. В. Налимов. – М. : Наука, 1974. – 272 с.

128. Нариньяни А. С. Кентавр по имени ТЕОН: Тезаурус + Онтология / А. С. Нариньяни // Межд. семинар Диалог 2001. – Аксаково. – Т. 1. – С. 199.
129. Нелюбин Л. Л. Перевод и прикладная лингвистика / Л. Л. Нелюбин. – М. : Высш. шк., 1983. – 207 с.
130. Никифоров А. Л. Семантическая концепция понимания / А. Л. Никифоров // Загадка человеческого понимания. – М. : Политиздат, 1991. – С. 72–94.
131. Нильсон Н. Искусственный интеллект. Методы поиска решения: Пер. с англ. / Н. Нильсон. – М. : Мир, 1973. – 271 с.
132. Новиков А. И. Семантика текста и ее формализация / А. И. Новиков. – М. : Наука, 1983. – 215 с.
133. Норман Д. Память и научение: Пер. с англ. / Д. Норман. – М. : Мир, 1985. – 160 с.
134. Орлова Л. В. Структура сверхфразового единства в научных текстах / Л. В. Орлова. – К. : Наукова думка, 1988. – 156 с.
135. Остин Дж. Значение слова / Дж. Остин // Аналитическая философия. Избранные тексты. – М. : Изд-во МГУ, 1993. – С. 105–121.
136. Осыка А. Ф. Обзор исследований по автоматической обработке естественного языка в США / А. Ф. Осыка // Проблемы бионики. – Харьков: Вища шк., 1985. – Вып. 34. – С. 3–18.
137. Павилёнис Р. И. Проблема смысла: современный логико-философский анализ языка / Р. И. Павилёнис. – М. : Мысль, 1983. – 286 с.
138. Партико З. В. Прикладна і комп'ютерна лінгвістика: Вступ до спеціальності: Навч. посіб. / З. В. Партико. – Львів: Афіша, 2008. – 224 с.
139. Пащенко Н. А. Проблемы автоматизации индексирования и реферирования / Н. А. Пащенко, Л. В. Кнорина, Т. В. Молчанова и др. // Итоги науки и техники. – М. : ВИНТИАН СССР, 1983. – Сер. Информатика. – Т. 7. – С. 7–164.
140. Перебийніс В. І. Статистичні методи для лінгвістів: Навчальний посібник / В. І. Перебийніс. – Вінниця: Нова Книга, 2001. – 168 с.

141. Перебейнос В. И. Морфологический анализ научного текста на ЭВМ / В. И. Перебейнос, Т. А. Грязнухина, Н. П. Дарчук и др. – К. : Наук. думка, 1989. – 264 с.

142. Перебийніс В. І. Коли робот заговорить (Мовне спілкування людини з електронно-обчислювальною машиною) / В. І. Перебийніс. – К. : Наукова думка, 1972. – 46 с.

143. Петров В. Н. Информационные системы / В. Н. Петров. – СПб.: Питер, 2003. – 688 с.

144. Пещак М. М. Лексична семантика в системі «людина-машина» / М. М. Пещак, Н. Ф. Клименко, Г. М. Ярун і др. – К. : Наукова думка, 1986. – 284 с.

145. Пиотровский Р. Г. Методы автоматического анализа и синтеза текста: Учеб. пособие для ин-тов и фак. иностр. яз. / Р. Г. Пиотровский, В. Н. Билан, М. Н. Боркун и др. – Мн. : Выш. шк., 1985. – 222 с.

146. Пиотровский Р. Г. Инженерная лингвистика и теория языка / Р. Г. Пиотровский. – Л. : Наука, 1979. – 112 с.

147. Пиотровский Р. Г. Математическая лингвистика: Учеб. пособие для пед. ин-тов / Р. Г. Пиотровский, К. Б. Бектаев, А. А. Пиотровская. – М. : Высш. шк., 1977. – 383 с.

148. Пиотровский Р. Г. Текст, машина, человек / Р. Г. Пиотровский. – Л. : Наука, 1975. – 327 с.

149. Пономаренко В. С. Проектування інформаційних систем: Посібник / В. С. Пономаренко. – К. : Академія, 2002. – 488 с.

150. Попов Э. В. Общение с ЭВМ на естественном языке / Э. В. Попов. – М. : Наука, 1982. – 360 с.

151. Попов Ю. В. Текст: структура и семантика / Ю. В. Попов, Т. П. Трегубович. – Минск: Выш. шк., 1984. – 190 с.

152. Попович М. В. Доказательство и понимание / М. В. Попович. – К. : Наукова думка, 1986. – С. 165.

153. Поспелов Д. А. Логико-лингвистические модели в системах управления / Д. А. Поспелов. – М. : Энергоиздат, 1981. – 232 с.
154. Правдин М. Н. Проблема абстрактного и конкретного в мышлении и языке / М. Н. Правдин // РАН. – М. : Вдохновение, 1991. – 229 с.
155. Приходько С. М. Автоматическое реферирование на основе анализа межфразовых связей / С. М. Приходько, Э. Ф. Скороходько // НТИ. – Сер. 2, 1982. – № 1. – С. 27–31.
156. Проблемно-ориентированные представления. – Владивосток: ИАПУ ДВНЦ АН СССР, 1985. – 27 с.
157. Ранганатан Ш. Р. Классификация двоеточием. Основная классификация: Пер. с англ. / Ш. Р. Ранганатан. – М. : Наука, 1970. – 423 с.
158. Ревзин И. И. Структура языка как моделирующей системы / И. И. Ревзин. – М. : Наука, 1978. – 288 с.
159. Романюк Ю. В. Граматичні процеси текстової компресії (на прикладі анафоричного еліпса) / Ю. В. Романюк. Дис... канд. філол. наук: 10.02.15. – К. : Наукова думка, 1996. – 21 с.
160. Рубашкин В. Ш. О методах анализа связного текста: к проблеме применения семантических моделей текста в документально-фактографических ИПС / В. Ш. Рубашкин // Вопросы информ. теории и практики. – Вып. 49. – 1983. – С. 58–73.
161. Рубашкин В. Ш. Представление и анализ смысла в интеллектуальных информационных системах / В. Ш. Рубашкин. – М. : Наука, 1989. – 192 с.
162. Русская грамматика. – М. : Наука, 1980. – Т. I, II. – 784 с., 710 с.
163. Рыбаков Ф. И. Автоматическое индексирование на естественном языке / Ф. И. Рыбаков, Е. А. Руднев, В. А. Петухов. – М. : Энергия, 1980. – 160 с.
164. Самойленко Т. Реферат и аннотация / Т. Самойленко, А. Сизенко // Секретарь-Референт (все о работе и для работы секретаря). – № 03 (28), 2005. – С. 40–41.

165. Севбо И. П. Структура связного текста и автоматизация реферирования / И. П. Севбо. – М. : Наука, 1969. – 136 с.
166. Скороходько Е. Ф. Лінгвістичні основи автоматизації інформаційного пошуку / Е. Ф. Скороходько. – К. : Вища школа, 1970. – 244 с.
167. Скороходько Е. Ф. Сіткове моделювання лексики: лінгвістична інтерпретація параметрів семантичної складності / Е. Ф. Скороходько // Мовознавство. – 1995. – № 6. – С. 19–28.
168. Скороходько Э. Ф. Лингвистические проблемы обработки текстов в автоматизированных информационно-поисковых системах / Э. Ф. Скороходько // Вопросы информационной теории и практики. – М. : ВИНТИ, 1974. – № 25. – С. 15–120.
169. Скороходько Э. Ф. Распределение межфразовых связей в тексте / Э. Ф. Скороходько // НТИ. – 1980. – Сер. 2. – № 4. – С. 23–30.
170. Скороходько Э. Ф. Семантические сети и автоматическая обработка текста / Э. Ф. Скороходько. – К. : Наукова думка, 1983. – 220 с.
171. Слобин Д. Психолінгвістика: Пер. с англ. / Д. Слобин, Дж. Грин. – М. : Прогресс, 1976. – 350 с.
172. Соловьев В. И. Подготовка и редактирование информационных изданий: Учеб. пособие / В. И. Соловьев. – М. : Изд-во МПИ, 1986. – 72 с.
173. Соловьев В. И. Реферат в научно-информационной деятельности и критерии его редакторской оценки / В. И. Соловьев. – Автореферат диссертации на соискание ученой степени кандидата филологических наук. – М. : Изд-во МПИ, 1970. – 17 с.
174. Соловьев В. И. Составление и редактирование рефератов: Вопросы теории и практики. Жанры информационной литературы: Обзор. Реферат / В. И. Соловьев. – М. : Книга, 1983. – 320 с.
175. Соловьев В. И. Поаспектный метод реферирования / В. И. Соловьев // НТИ. – Сер. 2. – 1971. – № 2. – С. 14–17.
176. Станислав Л. Сумма технологий / Л. Станислав. – М. – 1968. – 607 с.

177. Степанов Ю. С. Методы и принципы современной лингвистики / Ю. С. Степанов. – М. : Наука, 1975. – 312 с.
178. Стефанюк В. Л. Реальность и прогнозы искусственного интеллекта: Пер. с англ. / В. Л. Стефанюка. – М. : Мир, 1987. – 247 с.
179. Ступин В. С. Система автоматического реферирования методом симметричного реферирования / В. С. Ступин // Компьютерная лингвистика и интеллектуальные технологии. – М. : Наука, 2004. – С. 579–591.
180. Сэлтон Г. Автоматическая обработка, хранение и поиск информации: Пер. с англ. / Г. Сэлтон. – М. : Сов. радио, 1973. – 560 с.
181. Тулдава Ю. Проблемы и методы квантитативно-системного исследования лексики / Ю. Тулдава. – Таллин: Валгус, 1987. – 204 с.
182. Фафенбергер Б. Толковый словарь по компьютерным технологиям и Internet / Б. Фафенбергер, Д. Уолл. – К. : Диалектика, 1996. – 480 с.
183. Федосюк М. Ю. Неявные способы передачи информации в тексте: Учебное пособие по спецкурсу / М. Ю. Федосюк. – М. : МГПИ им. В. И. Ленина, 1988. – 83 с.
184. Фитиалов С. Я. Формальные грамматики: Учеб. пособие / С. Я. Фитиалов. – Л. : Наука, 1984. – 99 с.
185. Фланаган Джеймс Л. Анализ, синтез и восприятие речи: Пер. с англ. / Джеймс Л. Фланаган. – М. : Связь, 1968. – 396 с.
186. Хайрова Н. Ф. Автоматизированные информационные библиотечные системы: задачи обработки информации: Моногр. / Н. Ф. Хайрова, Н. В. Шаронова. – Харьков: Око, 2002. – 120 с.
187. Хайрова Н. Ф. Машинный перевод: Учеб. пособие для студентов лингв. спец. / Хайрова Н. Ф., Замаруева И. В. – Х. : Око, 1998. – 80 с.
188. Хайрова Н. Ф. Современные аспекты автоматического реферирования / Н. Ф. Хайрова // Ученые зап. Харьк. гуманит. ин-та «Нар. укр. акад.». – Харьков: Око, 1997. – Т. 3. – С. 364–374.

189. Хан У. Системы автоматического реферирования [Электрон. ресурс] / У. Хан, И. Мани / Сайт журнала «Открытые системы», 2000. – № 12. – С. 15–20. – Режим доступа: http://www.osp.ru/os/2000/12/067_print.htm.
190. Хант Э. Искусственный интеллект: Пер. с англ. / Э. Хант. – М. : Мир, 1978. – 559 с.
191. Хорафас Д. Н. Системы и моделирование: Пер. с англ. / Д. Н. Хорафас. – М. : Мир, 1967. – 420 с.
192. Цаленко М. Ш. Моделирование семантики в базах данных / М. Ш. Цаленко // Проблемы искусственного интеллекта. – М. : Наука, 1989. – 288 с.
193. Черняховская М. Ю. Представление знаний для диагностических экспертных систем в медицине / М. Ю. Черняховская. Владивосток: ДВО РАН СССР, 1983. – 179 с.
194. Чёрч А. Введение в математическую логику: Пер. с англ. / А. Чёрч. – М. : Изд-во ИЛ, 1960. – Т. 1. – 486 с.
195. Шабанов-Кушнаренко Ю. П. Компараторная идентификация лингвистических объектов: моногр. / Ю. П. Шабанов-Кушнаренко, Н. В. Шаронова. – К. : Наукова думка, 1993. – 116 с.
196. Шабанов-Кушнаренко Ю. П. Теория интеллекта. Математические средства / Ю. П. Шабанов-Кушнаренко. – Харьков: Вища шк., 1984. – 144 с.
197. Шавердова Л. Н. К вопросу об оптимальной структуре реферата / Л. Н. Шавердова // Реферирование в общественных науках. – М., 1982. – С. 101–114.
198. Шаронова Н. В. Направления совершенствования современных систем автоматического реферирования / Н. В. Шаронова, Н. Ф. Хайрова // Вест. Херсон. гос. техн. ун-та. – Херсон. – 1999. – № 1. – С. 78–80.
199. Шемакин Ю. И. Тезаурус в автоматизированных системах управления и обработки информации / Ю. И. Шемакин. – М. : Воениздат, 1974. – 192 с.

200. Шенк Р. Обработка концептуальной информации / Р. Шенк. – М. : Энергия, 1980. – 361 с.
201. Шиманов А. Е. Аннотирование и реферирование в издательском деле: Конспект лекций / А. Е. Шиманов. – М. : МГУП, 2004. – 131 с.
202. Шрейдер Ю. А. Семиотические основы информатики / Ю. А. Шрейдер. – М. : ИПКРИР. – 1974. – 81 с.
203. Ыйм Х. Эпизоды в структуре дискурса / Х. Ыйм // Представление знаний и моделирование процесса понимания. – Новосибирск: ВЦ СО АН СССР, 1980. – С. 79–96.
204. Якобсон Р. Избранные работы / Р. Якобсон. – М. : Прогресс, 1985. – 456 с.
205. Яковенко А. А. Модель синтаксического обобщения в процессе реферирования / А. А. Яковенко // Вестник МСУ. – 2002. – №3. – С. 17–19.
206. Яковенко А. А. Обобщение на морфологическом уровне в процессе реферирования / А. А. Яковенко, О. В. Лазаренко // Вестник МСУ. – 2001. – № 6. – С. 45–46.
207. Яковенко А. А. Проблема розуміння текстової інформації (прикладний аспект) / А. А. Яковенко // Проблеми семантики слова, речення та тексту. – К. : КНЛУ, 2001. – С. 223–229.
208. Яковенко А. А. Семантическое и синтаксическое обобщение в процессе реферирования / А. А. Яковенко // Вчені зап. Харк. гуманіт. ін-ту «Нар. укр. акад.». – Харьков: Око, 2001. – Т. 9. – С. 446–450.
209. Якушин Б. В. Алгоритмическое индексирование в информационных системах. Проблематика и методы / Б. В. Якушин. – М. : Наука, 1978. – 144 с.
210. Alberts L. K. YMIR: an Ontology for Engineering Design / L. K. Alberts. University of Twente, 1993. – 171 p.
211. Ando R. K. Multidocument Summarization by Visualizing Topical Content / R. K. Ando et al. // Proc. ANLP/NAACL Workshop on Automatic Summarization, 2000. – P. 79–88.

212. Benjamins V. R. Community is Knowledge [Электрон. ресурс] / V. R. Benjamins, D. Fensel // Сайт KAW'98, Eleventh Workshop on Knowledge Acquisition, Modeling and Management, Canada, 1998. – Режим доступа: <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/benjamins1>.

213. Bobrow D. G. An Overview of KRL, a Knowledge Representation Language / D. G. Bobrow, T. Winograd // Cognitive Science, 1977. – № 1. – P. 21–29.

214. Borko H. Abstracting concepts and methods / H. Borko, C. Bernier. – New-York: Academic Press, 1970. – 250 p.

215. Brandow R. Automatic condensation of electronic publications by sentence selection / R. Brandow, K. Mitze, L. Rau // In Information Processing & Management. – 1995. – № 31 (5). – P. 675–685.

216. Braslavski P. A User-Centered Comparison of Web Search Engines / P. Braslavski, A. Shishkin // In Computational Linguistics and Intelligent Technologies. – 2005. – P. 554–560.

217. Carbonell J. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries / J. Carbonell, J. Goldstein // Proc. 21st Int'l ACM SIGIR Conf. Research and Development in Information Retrieval. – New York: ACM Press, 1998. – P. 335–336.

218. Chandrasekaran B. The Ontology of Tasks and Methods [Электрон. ресурс] / B. Chandrasekaran, J. Josephson, V. Benjamins // Сайт KAW'98, Eleventh Workshop on Knowledge Acquisition, Modeling and Management, Canada, 1998. – Режим доступа: <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/chandra>.

219. Chomsky N. Syntactic structures / N. Chomsky. – The Hague: Mouton. – 1957.

220. Dijk van T. A. Semantic Macro-Structures and Knowledge Frames in Discourse Comprehension / T. A. van Dijk // Cognitive Processes in Comprehension. Hillsdale, N.J. : Lawrence Erlbaum, 1977. – P. 3–32.

221. Dubois J. Dictionnaire de linguistique et des science du langage / J. Dubois. – Paris, 1994. – 419 p.
222. ERGO Linguistic Technologies [Электрон. ресурс] / Сайт ERGO. – 2000. – Режим доступа: <http://www.ergo-ling.com>.
223. Extractor [Электрон. ресурс] / Сайт Института информационных технологий Национального исследовательского совета Канады. – 1998. – Режим доступа: http://ai.iit.nrc.ca/ll_public/extractor.html.
224. Fensel V. Ontobroker: Transforming the WWW into a Knowledge Base [Электрон. ресурс] / V. Fensel, S. Decker, M. Erdmann et al. // Сайт KAW'98, Eleventh Workshop on Knowledge Acquisition, Modeling and Management, 1998. – Режим доступа: <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/fensel1>.
225. Fikes R. Large-Scale Repositories of Highly Expressive Reusable Knowledge [Электрон. ресурс] / R. Fikes, A. Farquhar. – 1997. – Режим доступа: <http://www-ksl.stanford.edu/yearindex.html#1997>.
226. Frege G. On sense and reference / G. Frege // The logic of grammar. – Belmont, 1975. – P. 119.
227. Fridman N. Ontology design: a survey and comparative review / N. Fridman // AI Magazine. – 1997. – № 18 (3). – P. 53–74.
228. Fukusima T. Text Summarization Challenge – Text Summarization Evaluation in Japan / T. Fukusima, M. Okumura // Workshop on Automatic Summarization, North American Association for Computational Linguistics. – 2001. – P. 51–59.
229. Functional Dependency Grammar [Электрон. ресурс] / Сайт Conexor. – 1998. – Режим доступа: <http://www.conexor.fi>.
230. Genesereth M. R. Logical Foundation of Artificial Intelligence / M. R. Genesereth, N. J. Nilsson // Los Altos. – California: Morgan Kaufmann, 1987. – 217 p.
231. Google Search Appliance for Intranets [Электрон. ресурс] // Google Inc., 2004. – Режим доступа: http://www.google.com/appliance/pdf/ds_GSA_intranets.pdf.

232. Gruber T. R. A translation approach to portable ontology specifications / T. R. Gruber // Knowledge Acquisition. – 1993. – № 5. – P. 199–220.

233. Gruber T. R. Ontolingua: A Mechanism to Support Portable Ontologies / T. R. Gruber // Technical Report KSL-91-66. – Stanford University: Knowledge Systems Laboratory, 1992. – P. 1–28.

234. Gruber T. R. The role of common ontology in achieving sharable, reusable knowledge bases / T. R. Gruber // In Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference. – 1991. – P. 601–602.

235. Gruber T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing / T. R. Gruber // In Inter. Journal of Human-Computer Studies, 1994. – № 43 (5/6). – P. 907–928.

236. Guarino N. Understanding, Building and Using Ontologies / N. Guarino // IJHCS. – 1996. – № 46. – P. 2–3.

237. Guarino N. Formal Ontology and Information Systems / N. Guarino // In Proceeding of International Conference on Formal Ontology in Information Systems (FOIS'98). – Amsterdam: IOS Press, 1998. – P. 3–15.

238. Guarino N. OntoSeek: Content-Based Access to the Web / N. Guarino // IEEE Intelligent Systems. – 1999. – P. 70–80.

239. Guarino N. The role of ontology in the information technology / N. Guarino, R. Poli // International Journal of Human-Computer Studies. Special issue on ontology. – 1995. – № 43 (5/6). – P. 623–965.

240. Guiraud P. La semantique / P. Guiraud. – Paris, 1955. – P. 15.

241. Hahn U. The Challenges of Automatic Summarization / U. Hahn, I. Mani // IEEE Computer. – 2000. – P. 29–36.

242. Hahn U. Knowledge-Based Text Summarization: Saliency and Generalization Operators for Knowledge-Based Abstraction / U. Hahn, U. Reimer // Advances in Automatic Text Summarization. – Cambridge: MIT Press, 1999. – P. 215–232.

243. Hayes-Roth F. The State of Knowledge-Based Systems / F. Hayes-Roth, N. Jacobstein // Communications of the ACM. – 1994. – V. 37. – № 3. – P. 27–39.

244. Heijst G. Using Explicit Ontologies in KBS Development / G. Heijst, A. Schreiber, B. Wielinga // In International Journal of Human and Computer Studies. – 1996. – № 46 (2/3). – P. 183–292.

245. Hopfield J. J. Neural networks and physical systems with emergent collective computational abilities / J. J. Hopfield // Proc. Nat. Acad. Sci. USA. – 1982. – P. 2554–2552.

246. Human Language Technology Sector of the Information Society Technologies (IST) Programme 1998 – 2000 [Электрон. ресурс] / Сайт отдела естественно-языковых технологий Общества информационных технологий. – 2000. – Режим доступа: <http://www.linglink.lu>.

247. Hutchins J. Summarization: Some Problems and Methods / J. Hutchins // Proc. Informatics 9: Meaning-The Frontier of Informatics. – London, 1987. – P. 151–173.

248. Iatsko V. Linguistic Aspects of Summarization [Электрон. ресурс] / V. Iatsko // In Philologieim Netz. – Vol. 18. – 2001. – P. 33-46. – Режим доступа: <http://www.fuberlin.de/phin/phin18/p18t3.htm>.

249. Inxight Summarizer [Электрон. ресурс] / Сайт Inxight. – 1995. – Режим доступа: <http://www.inxight.com>.

250. Khoroshevsky V. F. Knowledge v. s. data spaces: how an applied semiotics to work on web / V. F. Khoroshevsky // Proceedings of CAI'98. – Pushchino (Russia). – 1998. – P. 7–16.

251. Krvavac A. Procedure in Abstract writing / A. Krvavac // Termnet News. – 1987. – № 19. – P. 5-8.

252. Kupiec J. A trainable document summarizer / J. Kupiec, J. Pedersen, C. Francine // Proc.18th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval. – New York: ACM Press. – 1995. – P. 68–73.

253. Lee S. Ontology-based Knowledge Discovery on the World-Wide-Web [Электрон. ресурс] / S. Lee, D. Rager // Proceedings of the Workshop on Internet-

based Information Systems, AAAI-96. – Oregon, 1996, – Режим доступа: <http://www.cs.umd.edu/projects/plus/SHOE/pubs/aaai-paper.html>.

254. Lenat D. B. CYC: a large-scale investment in knowledge infrastructure / D. B. Lenat // Communications of the ACM 38. – 1995. – № 11. – Режим доступа: <http://www.cyc.com>.

255. Lerat P. Semantique descriptive / P. Lerat. – Paris, 1983. – P. 37.

256. Lienhart R. Video Abstracting / R. Lienhart, S. Pfeiffer, W. Effelsberg // Comm. ACM, 1997. – Vol. 40. – № 12. – P. 54–62.

257. Lin C. Manual and Automatic Evaluation of Summaries / C. Lin, E. Hovy // Proc. of the Workshop on Automatic Summarization (including DUC 2002), Association for Computational Linguistics. – 2002. – P. 45–51.

258. Link Grammar Homepage [Электрон. ресурс] / Сайт Link. – 2001. – Режим доступа: <http://bobo.link.cs.cmu.edu/link>.

259. Luhn H. P. The automatic creation of literature abstracts / H. P. Luhn // Advances in automatic text summarization. – The MIT Press, 1999. – P. 15–21.

260. Luke S. Applying Ontology to the Web: A Case Study [Электрон. ресурс] / S. Luke, J. Helfin, J. Hendler. – 1999. – Режим доступа: <http://www.cs.umd.edu/projects/plus/SHOE/iwann99.pdf>.

261. Lyons J. Elements de semantique / J. Lyons. – Paris, 1978. – P. 9.

262. Maizell R. E. Abstracting scientific and technical literature / R. E. Maizell et al. – New-York: Willey, 1971. – 297 p.

263. Mani I. Advances in automatic text summarization / I. Mani, M. Maybury. – London: The MIT Press Cambridge, 1999. – 434 p.

264. Mani I. Summarizing Similarities and Differences Among Related Documents / I. Mani, E. Bloedorn // Information Retrieval. – Vol. 1. – № 1. – 1999. – P. 35–67.

265. Mani I. The Tipster Sumac Text Summarization Evaluation / I. Mani et al. // Proc. 9th Conf. European Chapter of the November, 2000. – P. 37–42.

266. Mark B. Ontologies as the Representation (and Re-Representation) of Agreement [Электрон. ресурс] / B. Mark. – Режим доступа: <http://www-ksl.stanford.edu/KR96/BillMarkPositionPaper.html>.

267. Merlino A. Broadcast News Navigation Using Story Segments / A. Merlino, D. Morey, M. Maybury // ACM Multimedia 97. – New York: ACM Press, 1997. – P. 381–391.

268. Michalski R. Learning from Observation: Conceptual Clustering / R. Michalski, R. Stepp // In Machine Learning: An Artificial Intelligence Approach. – California: Tioga, 1983. – P. 332.

269. Mizoguchi R. A Step Towards Ontological Engineering [Электрон. ресурс] / R. A. Mizoguchi. – Режим доступа: <http://www.ei.sanken.osaka-u.ac.jp/english/step-onteng.html>.

270. Nomoto T. The diversity-based approach to open-domain text summarization / T. Nomoto, Y. Matsumoto // In Information Processing & Management, 2003. – № 39. – P. 363–389.

271. OWL, язык веб-онтологий. Руководство. Режим доступа: <http://www.w3.org/TR/2004/REC-owl-guide-20040210>.

272. Peter S. A Practical Approach to Achieve Knowledge Sharing and Reuse / S. Peter, D. McGonigal // In Appl. of Advanced Information Technologies: Proceeding of the Forth World Congress on Expert systems. – Mexico: City Cognizant Communication Corporation, 1998. – Vol.1. – P. 434–441.

273. Prosum Summarizer [Электрон. ресурс] / Сайт Prosum. – 1999. – Режим доступа: <http://transend.labs.bt.com/cgi-bin/prosum/prosum>.

274. Radev D. Evaluation challenges in large-scale document summarization / D. Radev, W. Lam, C. Arda et al. // Proc. of the 41st Annual Meeting of the Association for Computational Linguistics. – 2003. – P. 375–382.

275. Radev D. Introduction to the special issue on summarization / D. Radev, K. McKeown, E. Hovy // Computational Linguistics. – 2002. – Vol. 28. – № 4. – P. 399–408.

276. Radev D. R. Generating Natural Language Summaries from Multiple Online Sources / D. R. Radev, K. R. McKeown // *Computational Linguistics*. – 1998. – Vol. 24. – № 3. – P. 469–500.

277. Salton G. Automatic Text Structuring and Summarization / G. Salton et al. // *Information Processing & Management*. – 1997. – Vol. 33. – № 2. – P.193–207.

278. Schreiber A. T. CML: The Common KADS Conceptual Modeling Language / A. T. Schreiber, B. J. Wielinga, J.M. Akkermans // *In A Future for Knowledge Acquisition. Proceedings of the 8th European Knowledge Acquisition Workshop EKAW*. – Berlin/Heidelberg: Springer-Verlag, 1995. – P. 1–25.

279. Smith K. E. Hypertext – linking to the future / K. E. Smith // *Online*. – 1988. – Vol. 12. – № 2. – P. 32–40.

280. Sowa J. Conceptual Structures: Information Processing in Mind and Machine / J. Sowa // *Reading*. – MA: Addison-Wesley, 1984. – P. 16.

281. Sowa J. F. Knowledge Representation: Logical, Philosophical and Computational Foundations / J. F. Sowa. – California: Brooks Cole Publishing C°, Pacific Grove, 2000. – 117 p.

282. Stepp R. Conceptual Clustering of Structured Objects: A Goal-Oriented Approach / R. Stepp, R. Michalski // *Artificial Intelligence*. – 1986. – Vol. 28 (1). – P. 53.

283. Studer R. Knowledge Engineering: Principles and Methods / R. Studer, V. Benjamins, D. Fensel // *In Data & Knowledge Engineering*. – 1998. – № 25. – P. 161–197.

284. Takeda H. Collaborative development and Use of Ontologies for Design / H. Takeda, M. Takaai, T. Nishida // *Proceedings of the Tenth International IFIP WG 5.2/5.3 Conference PROLAMAT*. – Italy, 1998. – P. 51–62.

285. Teufel S. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status / S. Teufel, M. Moens // *Computational Linguistics*. – 2002. – Vol. 28. – № 4. – P. 409–445.

286. TextAnalyst [Электрон. ресурс] / Сайт «МикроСистемы». – 2000. – Режим доступа: <http://www.host.ru/~analyst>.
287. Tombros A. Advantages of Query Biased Summaries in Information Retrieval / A. Tombros, M. Sanderson // In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. – Melbourne (Australia). – 1998. – P. 2–10.
288. TOVE Manual // Department of Industrial Engineering. – Toronto: University of Toronto, 1999. – Режим доступа: <http://www.ie/utoronto.ca/EIL/tove>.
289. Uschold M. Knowledge Level Modeling: Concepts and Terminology / M. Uschold // In The Knowledge Engineering Review. – 1998. – Vol. 13 (1). – P. 5–9.
290. Ushold M. Ontologies: Principles, Methods and Applications / M. Ushold, M. Gruninger // Knowledge Engineering Review. – 1996. – Vol. 11. – № 2. – P. 11–17.
291. Ushold M. Towards a Methodology for Building Ontologies / M. Ushold, M. King // Workshop on Basic Ontologica Issues in Knowledge Sharing. – 1995. – P. 41–49.
292. Vet P. E. The Plinius Ontology of Ceramic Materials / P. E. Vet, N. J. Mars, P. H. Speel // ECAI. – Amsterdam, 1994. – P. 8–12.
293. Wasson M. Using Summaries in Document Retrieval / M. Wasson // Proc. of the Workshop on Automatic Summarization. – Association for Computational Linguistics. – 2002. – P. 37–44.
294. Weil B. H. Standards for writing abstracts/ B. H. Weil // J. Amer. Soc. Inf. Sci. – 1970. – Vol. 21. – № 5. – P. 27–84.
295. Weiss S. A Model-Based Method for Computer-Aided Medical Decision Making / S. Weiss, C. Kulikovski et al. // In Reading in Medical Artificial Intelligence. – Addison Wesley, 1984. – P. 61–67.
296. White R. W. A task-oriented study on the influencing effects of query-biased summarization in web searching / R. W. White, J. M. Jose, I. Ruthven // In Information Processing & Management. – 2003. – № 39. – P. 707–733.

297. Wielinga B. J. Reusable and Sharable Knowledge Bases: a European Perspective / B. J. Wielinga, A. T. Schreiber // In Proceedings International Conference on Building and Sharing of Very Large-Scaled Knowledge Bases. – Tokyo: Japan Information Processing Development Center, 1993. – P. 103-115.

298. Wielinga B. Framework and Formalism for Expressing Ontologies (Version 1) / B. Wielinga, A. T. Schreiber, W. Jansweijer et al. // ESPRIT Project 8145 KACTUS. – Free University of Amsterdam Deliverable, DO1b.1. – 1994. – P. 1–15.

299. Wierzbicka A. Semantic primitives / A. Wierzbicka. – Frankfurt, 1972. – 235 p.

300. Wyllys R. E. Extracting and Abstracting by Computer / R. E. Wyllys // Automated Language Processing. – New York, 1967. – P. 127–179.

301. Yang C. C. Fractal Summarization for Mobile Devices to Access Large Documents on the Web / C. C. Yang, F. L. Wang // In Proceedings of the WWW, Budapest. – 2003. – Режим доступа: <http://www2003.org/cdrom/papers/refereed/p681/p681-yanghtml/p681-yang.html>.

302. Zhou L. A Web-Trained Extraction Summarization System / L. Zhou, E. Hovy // Proc. of HLT-NAACL. – 2003. – P. 205–211.

**Приклади представлення інтелектуальних індикативних рефератів в
уніфікованому вигляді**

<i>Вихідний реферат</i>	<i>Уніфікований реферат</i>
<p>Рассматриваются вопросы усовершенствования финансово-кредитного механизма АПК. Обращается внимание на необходимость формирования системы ипотечного и товарного кредитования и сезонного кредитования под залог сельскохозяйственной продукции. Отмечена необходимость организации финансового консалтинга и развития лизинга сельскохозяйственной техники, оборудования и технологий.</p>	<p>Рассматриваются вопросы усовершенствования финансово-кредитного механизма АПК. Анализируется необходимость формирования системы ипотечного, товарного кредитования и сезонного кредитования. Отмечается необходимость организации финансового консалтинга и развития лизинга сельскохозяйственной техники, оборудования и технологий.</p>
<p>В статье анализируется инвестиционная привлекательность различных регионов Украины в контексте современных экономических преобразований. В результате проведенного исследования формулируется система практических рекомендаций для более успешного привлечения потенциальных инвесторов.</p>	<p>Анализируется инвестиционная привлекательность различных регионов Украины. Формулируется система рекомендаций для привлечения потенциальных инвесторов.</p>
<p>Проведен анализ принадлежности различных соционических психотипов к темпераментам по типологии Хейманса-Ле Сенна. Приведены результаты практической работы, из которых следует, что внутри представителей каждого психотипа встречается как минимум два темперамента.</p>	<p>Проанализированы соционические психотипы по типологии Хейманса-Ле Сенна. Приведены результаты практической работы.</p>

<p>Рассматривается системный подход к обоснованию и моделированию социетальной психики неслучайных групп. Показано, что с абстрактной точки зрения вполне правомерно представление структуры и механизма функционирования социетальной психики неслучайных групп теми же моделями, что и психики личности. Во второй части статьи в качестве примера рассмотрен интегральный тип информационного метаболизма нации.</p>	<p>Излагается проблема обоснования и моделирования социетальной психики неслучайных групп. Доказывается правомерность представления структуры и механизма функционирования психики групп моделями психики личности. Рассматривается интегральный тип информационного метаболизма нации.</p>
<p>Рассмотрено соответствие специфики медицинских специальностей и требований к ним и характеристик соционических типов личности. Предложены конкретные методические рекомендации по осуществлению профессионально-психологического отбора и распределения людей определенных типов по врачебным специальностям.</p>	<p>Проанализированы соционические типы личности медицинских специальностей. Предложены рекомендации профессионально-психологического отбора для распределения людей по врачебным специальностям.</p>
<p>Сформулированы основные положения интегральной соционики, рассматриваются интегральные типы информационного метаболизма коллективов, организаций, наций, этносов и государств, их психоинформационная структура и взаимодействия. В рамках этносоционики как составной части интегральной соционики рассмотрены методы описания типов этносов и государств.</p>	<p>Сформулированы основные положения интегральной соционики. Рассмотрены интегральные типы информационного метаболизма коллективов, организаций, наций, этносов и государств, их психоинформационная структура и взаимодействия. Даны этносоционические методы описания типов этносов и государств.</p>
<p>Методами этносоционики дано покомпонентное описание и анализ модели интегрального типа США как ЛИЭ. Рассмотрены психоинформационные взаимодействия США с другими</p>	<p>Методами этносоционики описана и проанализирована модель интегрального типа США. Рассмотрены психоинформационные взаимодействия США.</p>

государствами и этносами.	
<p>Освещено современное состояние проблемы лечения рака мочевого пузыря. Рассмотрена эффективность применяемых методов химиолучевого хирургического лечения, иммунотерапии. Сделан вывод о необходимости дальнейших исследований эффективности адьювантной, неоадьювантной химиотерапии и комбинированного паллиативного и симптоматического лечения рака мочевого пузыря.</p>	<p>Освещена проблема лечения рака мочевого пузыря. Рассмотрена эффективность методов химиолучевого хирургического лечения, иммунотерапии. Показана необходимость дальнейших исследований эффективности химиотерапии и комбинированного лечения рака мочевого пузыря.</p>
<p>Представлено современное определение нестабильной стенокардии, которая объединяет группу клинических форм ИБС, характеризующихся повышенной вероятностью развития острого инфаркта миокарда. Приведены клинические и электрокардиографические признаки дестабилизации течения ИБС, патогенетические и клинические особенности различных форм нестабильной стенокардии. Представлены методы купирования ангинозного приступа и неотложные мероприятия при нестабильной стенокардии в зависимости от гемодинамических особенностей действия препаратов и ведущего патогенетического механизма в формировании определенных клинических форм ИБС.</p>	<p>Представлено современное определение нестабильной стенокардии. Приведены признаки дестабилизации ИБС. Рассмотрены особенности различных форм нестабильной стенокардии. Представлены методы купирования ангинозного приступа и неотложные мероприятия при нестабильной стенокардии.</p>
<p>Рассмотрены основные аспекты пересадки печени, подробно описаны особенности отбора доноров и реципиентов, этапы хирургического вмешательства, анестезиологическое пособие,</p>	<p>Рассмотрены основные аспекты пересадки печени. Подробно описаны особенности отбора доноров и реципиентов, этапы хирургического вмешательства, анестезиологическое пособие,</p>

<p>возможные осложнения. Показаны дальнейшие перспективы трансплантации печени, актуальность и значимость проблемы.</p>	<p>возможные осложнения. Показаны актуальность и значимость дальнейшей перспективы трансплантации печени.</p>
<p>Предлагается оригинальный подход к построению описания процесса понимания естественно-языкового текста, основанный на стандартизации семантической структуры предметной области с помощью предикатно-аргументных выражений с эксплицированными семантико-синтаксическими отношениями.</p>	<p>Предлагается оригинальный способ описания процесса понимания естественно-языкового текста.</p>
<p>Дана токсиколого-гигиеническая и медико-биологическая характеристика поверхностно-активных веществ, широко используемых в производстве и быту. Обобщены данные литературы и результаты собственных исследований медико-биологических аспектов проблемы загрязнения окружающей среды поверхностно-активными веществами.</p>	<p>Рассматриваются поверхностно-активные вещества. Обобщена проблема загрязнения окружающей среды поверхностно-активными веществами.</p>
<p>Предложен обзор материалов, опубликованных в США по проблеме обработки с помощью ЭВМ информации на языке человека. Особое внимание уделено методам перехода от текста на естественном языке к представлению его значения на языке машинной системы, методам построения лингвистического интерфейса для банков данных, анализу результатов практического использования построенных систем.</p>	<p>Предложен обзор проблемы машинной обработки естественно-языковой информации. Рассмотрены методы построения лингвистического интерфейса банков данных. Проанализированы результаты практического использования построенных систем.</p>
<p>Описывается общая структура языка запросов для системы управления базой данных. Установлены группы работающих с</p>	<p>Описывается структура языка запросов системы управления базой данных. Установлены группы пользователей. Дана</p>

<p>системой пользователей и дана характеристика выполняемых ими функций.</p>	<p>характеристика функций пользователей.</p>
<p>Рассматривается возможный подход к организации анализа естественно-языковых высказываний в диалоговых системах общения. Подход предполагает параллельное использование процедур морфологического, синтаксического, прагматического и информационного анализов для скорейшего достижения конечных результатов анализа.</p>	<p>Рассматривается организация анализа естественно-языковых высказываний в диалоговых системах общения. Описаны методы скорейшего достижения конечных результатов анализа.</p>
<p>Обсуждаются основные принципы формирования финансовых ресурсов региона. Рассматриваются источники формирования доходной части регионального бюджета. Особо отмечается зависимость структуры финансовых ресурсов от уровня концентрации производственных объектов на территории региона.</p>	<p>Обсуждаются основные принципы формирования финансовых ресурсов региона. Рассматриваются источники формирования доходной части регионального бюджета. Особо отмечается зависимость структуры финансовых ресурсов от уровня концентрации производственных объектов на территории региона.</p>
<p>Рассматривается необходимость математического описания смысла, выражаемого на естественном языке, предлагается новая наука о смысле. Обоснована важность изучения смысла на словообразовательном уровне. Исследуется структура словообразовательных значений глаголов русского языка.</p>	<p>Рассматривается необходимость математического описания смысла. Предлагается новая наука о смысле. Исследуется структура словообразовательных значений глаголов русского языка.</p>
<p>Отражены существующие подходы к реструктуризации железных дорог мира и возможности перенесения зарубежного опыта в Украину.</p>	<p>Рассмотрены возможности использования зарубежного опыта реструктуризации железных дорог в Украину.</p>

Варианти використання синтаксичної конструкції

$Vf_{3s(pl)}-N_1$ у реферативних текстах

1. $\{Vf_{3pl}\}-\{N_1\}-\{N_2\}-\{Adj\}-\{N_2\}-\{N_2\}$:

Рассматриваются вопросы усовершенствования финансово-кредитного механизма АПК;

2. $\{Vf_{3pl}\}-\{N_1\}-\{N_2\}-\{Adj\}-\{N_2\}-\{Adj\}-\{N_2\}$:

Рассматриваются источники формирования доходной части регионального бюджета;

3. $\{Vf_{3s}\}-\{N_1\}-\{Adj\}-\{N_2\}-\{N_2\}$:

Рассматривается необходимость математического описания смысла;

4. $\{Vf_{3pl}\}-\{N_1\}-\{Adj\}-\{N_2\}-\{N_2\}-\{Adj\}-\{N_2\}$:

Исследуется структура словообразовательных значений глаголов русского языка;

5. $\{Vf_{3s}\}-\{Adj\}-\{N_1\}-\{Adj\}-\{N_2\}-\{N_2\}$:

Анализируется инвестиционная привлекательность различных регионов Украины;

6. $\{Vf_{3pl}\}-\{Adj\}-\{N_1\}-\{N_2\}-\{Adj\}-\{N_2\}-\{N_2\}$:

Обсуждаются основные принципы формирования финансовых ресурсов региона;

7. $\{Vf_{3s}\}-\{N_1\}-\{Adj\}-\{N_2\}-\{Adj\}-\{N_2\}-\{Adj\}-\{N_2\}$:

Формулируется система практических рекомендаций для более успешного привлечения потенциальных инвесторов;

8. $\{Vf_{3pl}\}-\{Adj\}-\{N_1\}-\{Adj\}-\{N_2\}-\{N_2\}-\{N_2\}-\{N_2\}-\{N_2\}-\{N_2\}-\{Adj\}-\{N_1\}-\{N_1\}$:

Рассматриваются интегральные типы информационного метаболизма коллективов, организаций, наций, этносов и государств, их психоинформационная структура и взаимодействия;

9. {Vf_{3s}}-{Adj}-{N₁}-{N₆}:

Предлагается новая наука о смысле;

10. {Vf_{3s}}-{Adj}-{N₁}-{N₂}-{N₂}-{N₂}-{N₂}-{N₅}-{N₂}:

Описывается общая структура языка запросов для системы управления базой данных;

11. {Vf_{3s}}-{Adj}-{N₁}-{N₂}-{N₂}-{Adj}-{N₂}-{Adj}-{N₄}-{Adj}-{N₄}-{N₄}:

Предлагается концептуальная схема взаимосвязи предприятия и его внешнего окружения через информационные поля, информационные образы и знания;

12. {Adv}-{Vf_{3s}}-{N₁}-{N₂}-{Adj}-{N₂}-{N₂}-{N₂}-{Adj}-{N₂}:

Особо отмечается зависимость структуры финансовых ресурсов от уровня концентрации производственных объектов;

13. {Vf_{3s}}-{N₁}-{N₄}-{N₂}-{N₂}-{Adj}-{Adj}-{N₂}-{Adj}-{N₂}:

Обращается внимание на необходимость формирования системы ипотечного и товарного кредитования и сезонного кредитования;

14. {Vf_{3s}}-{Adj}-{N₁}-{N₃}-{N₃}-{Adj}-{N₂}-{Adj}-{N₂}:

Рассматривается системный подход к обоснованию и моделированию социальной психики неслучайных групп;

15. {Vf_{3s}}-{Adj}-{N₁}-{N₃}-{N₂}-{N₂}-{N₂}-{Adj}-{N₂}:

Предлагается оригинальный подход к построению описания процесса понимания естественно-языкового текста;

16. {Vf_{3s}}-{Adj}-{N₁}-{N₃}-{N₂}-{Adj}-{N₂}-{Adj}-{N₆}-{N₂}:

Рассматривается возможный подход к организации анализа естественно-языковых высказываний в диалоговых системах общения.

Загальнонаукова лексика

<i>m₄</i>	<i>m₅</i>	<i>m₆</i>
актуальность важность возможность значимость необходимость особенность правомерность эффективность специфика	анализ алгоритм аспект взаимосвязь вопрос данные зависимость знание источник материал метод механизм минимум наука обзор область объект определение перспектива подход положение признак принцип причина проблема процесс результат рекомендация свойство система соответствие состояние специальность средство структура схема техника технология форма функция характеристика	действие достижение изучение использование исследование обоснование обработка описание организация осуществление построение представление преобразование развитие распределение распространение стандартизация усовершенствование формирование

замечание, замешательство, замысел, запас, записка, запись, запоздание, запрос, запуск, заработок, заслуга, застой, затруднение, зачаток, заявка, заявление, звено, знание, значение, значимость, значительность, зона, зрелость, идеал, идеализм, идеология, идея, избыток, издание, изделие, изменение, изобретение, изученность, иллюзия, имущество, инвентарь, индекс, индивид, индивидуализм, индивидуум, индикатор, индукция, индустриализация, инерция, инициатива, институт, инструкция, инструмент, интеграция, интерполяция, интерпретация, информация, иррационализм, исключение, исключительность, искусство, исправление, испытание, исследование, исследователь, истина, источник, исход, итог, капания, канал, кандидат, кандидатура, канон, картотека, каталог, категория, каузальность, кафедра, качество, квалификация, квинтэссенция, кибернетизация, класс, классификатор, классификация, клиент, клиентура, книга, код, кодекс, колебание, количество, коллегия, коллектив, колорит, комиссия, компетенция, компиляция, комплекс, комплект, компонент, компоновка, конгломерат, конкурент, конкуренция, консерватизм, консолидация, конспект, константа, консультация, консультант, контраст, контроль, конференция, конфронтация, концентрация, концепция, конъюнктура, кооперация, координированность, копия, корпорация, корректив, корректировка, корреляция, коэффициент, критика, кульминация, культура, лаборатория, лектор, лекция, либерализм, литература, ликвидация, лимит, лимитация, литература, личность, логика, льгота, магистр, магистратура, макет, максимум, мнение, манипуляция, манускрипт, марка, масса, мастер, мастерство, масштаб, материал, материя, машина, метод, методика, методология, механизм, минимум, минус, мировоззрение, миропонимание, мобилизация, модель, модификация, монитор, мораль, мотив, мутация, мысль, наблюдение, набор, надпись, название, назначение, наименование, накопление, наличие, намерение, направление, наука, начинание, невозможность, невыполнение, недостаток, незавершенность, незнание, неисправность, необходимость, неосуществимость, непонимание, непризнание, неравенство, несоблюдение, несогласие, несоответствие, неспособность, несходство, неточность, неясность, новаторство, новация, новообразование, новшество, номенклатура, норма, норматив, нотация, нравственность, нумерация, нюанс, обеспечение, обеспеченность, обзор, обобщение, обозначение, обозрение, обоснование, образец, образование, обращение, обстановка, обстоятельство, обусловленность, общественность, общество, общность, объект, объективность, объяснение, оговорка, ограничение, одобрение, окончание, оператор, операция, описание, оппозиционность, оппозиция, оппонент, определение, определитель, опровержение, опыт, организация, оригинал, ориентация, осложнение, оснащение, основа, основание, особенность, остаток, ответственность, отзыв, отклик, отклонение, открытие, отличие, отметка, отношение, отрасль, отрицание, отсутствие, отчет, отчетность, очертания, ошибка, пакет, памятка, парадокс, параллель, патология, первенство, первоисточник, переговоры, перегрузка, период, персонал, перспектива, план, плоскость, плюс, поведение, повреждение, подготовленность, подобие, подотчетность, подразделение, подтверждение, подтекст, подход, подъем, познание, поиск, показатель, полемика, политика,

полномочие, полярность, понимание, поправка, порицание, последствие, поставка, постановление, построение, потенциал, потенция, потеря, поток, потребитель, потребность, правдоподобие, правление, право, прагматизм, практик, практика, превосходство, предисловие, предмет, предназначение, предосторожность, предписание, предположение, предпосылка, предприятие, председатель, представитель, представление, предубеждение, предупреждение, президент, презумпция, преимущество, прейскурант, премия, преобразование, препятствие, прерогатива, преступление, претендент, претензия, прецедент, прибавка, прибыль, привилегия, прием, приложение, применимость, применяемость, пример, примитив, примитивизм, принадлежность, принцип, приобретение, приоритет, прирост, приспособление, притязание, причина, проблема, проблематика, прогноз, программа, прогресс, продукт, продукция, проект, производство, производительность, производство, промежуток, промышленность, прообраз, пропаганда, пространство, протест, противовес, противодействие, противоположность, противоречие, протокол, протяженность, профанация, профессионал, профессионализм, профессия, профилактика, профорганизация, профориентация, профсоюз, процент, процесс, проявление, публикация, пункт, работа, равенство, равновесие, равноправие, радикализм, разбирательство, развитие, раздел, размер, размещение, разница, разновидность, разногласие, разнообразие, разработанность, разработка, разрешение, разрушение, разряд, разум, район, ранг, рапорт, расписание, расположение, распорядок, распоряжение, распределение, рассказ, расстояние, рассуждение, расхождение, расчет, ратификация, рационализация, реакция, реалия, реальность, ревизия, революция, регион, регистр, регламент, регресс, редакция, режим, резерв, резолюция, резонанс, результат, реклама, рекламация, рекомендация, рекорд, реорганизация, репрезентация, реферат, реформа, рецензия, решение, реэкспорт, риск, род, роль, рубеж, рубрика, руководитель, руководство, рукопись, ручательство, рынок, ряд, саботаж, самоконтроль, самокритика, самообразование, самообучение, самоокупаемость, самоподготовка, самопроверка, самосовершенствование, самоучитель, сбой, сведение, свидетельство, свобода, своеобразие, свойство, связь, сдвиг, сделка, сеанс, себестоимость, сезонность, секретарь, сектор, секция, сенсация, сентенция, сепаратор, сепарация, сердцевина, серийность, серия, сертификат, сеть, сигнал, сила, символ, символизация, символизм, символика, симметрия, симпозиум, симптом, симуляция, синкретизм, синопсис, синтез, синхронизм, синхрония, синхронность, система, систематизация, систематика, ситуация, скачок, скелет, скептицизм, склад, склонность, скорость, слава, словарь, слой, случай, случайность, слышимость, смена, сменяемость, смертность, смесь, смысл, снабжение, снаряд, снаряжение, соавтор, соавторство, собеседование, собрание, собственность, событие, совершенство, совет, совещание, совладелец, совладение, совместимость, совокупность, совпадение, современник, современность, согласие, соглашение, содержание, содержимое, содружество, соединение, создание, создатель, сознание, сознательность, соискательство, солидарность, сомнение, соображение, сообщение, сообщество, соответствие, соотношение, сопоставимость, сопричастность, соразмерность, сорт, сортность,

сосредоточенность, состав, состояние, состоятельность, сотрудничество, софизм, софистика, сочлен, союз, спектр, спекуляция, специалист, специальность, специфика, спецификация, спецкурс, список, спор, способность, спрос, сработанность, средство, срок, ссылка, стабилизация, стадия, стадность, стажировка, стандарт, стандартизация, становление, станок, статистика, статус, степень, стереотип, стечение, стилизация, стиль, стимул, стимулятор, стоимость, стоицизм, столкновение, сторона, стратегия, стратификация, страхование, строение, строительство, строй, структура, структурность, студент, студия, ступень, стык, субординация, субсидия, субстанция, субституция, субстрат, субъект, субъективизм, суверенитет, суд, судьба, суждение, сумма, суррогат, существо, существование, сущность, сфера, схема, схематизм, сходство, схоластика, счет, табель, таблица, такса, таксация, тактика, тариф, тарификация, творение, творчество, теза, тезаурус, тезис, текст, тело, тема, тематика, темп, темперамент, температура, тенденциозность, тенденция, теорема, теория, термин, терминология, территория, тест, техника, технология, тип, типаж, типизация, типология, товар, тождество, толкование, торговля, традиция, транспорт, трансформация, трафарет, требование, тренинг, тройственность, труд, убеждение, убыток, угроза, указание, уклад, умение, умозаключение, умозрение, универсализм, университет, упадок, уровень, условие, устройство, утилизация, участие, участок, учение, ущерб, фабрика, фаза, факт, фактичность, фактор, фактура, факультет, фальсификат, фальсификация, феномен, фигура, фикция, финал, фирма, фонд, форма, формализм, формация, формула, фрагмент, фракция, функция, халатность, характер, характеристика, ход, хозяйство, хроника, хронология, хронометраж, целенаправленность, цель, цена, цензура, ценность, центр, церемония, цивилизация, цикл, цитата, часть, человек, человечество, черта, численность, число, член, шаблон, шаг, ширина, шифр, шкала, штамп, штат, эвакуация, эволюция, эвристика, эгоцентризм, эквивалент, экзамен, экземпляр, эклектизм, эклектика, экономика, экономия, экспансия, эксперимент, экспертиза, экспликация, эксплуатация, экспозиция, экспорт, экстраполяция, экстремизм, эксцесс, элемент, элиминация, эмансипация, эмпиризм, энергия, энциклопедизм, эпизод, эпилог, эпоха, эрудиция, эталон, этап, эффект, юстиция, явление, ядро };

Результат

$N(m_5, ls_2)$ = { абстрагирование, автоматизация, агитация, адаптация, администрирование, адсорбция, акклиматизирование, аккомпанирование, аккумуляирование, активизация, активирование, акцентирование, амальгамирование, американизация, анализирование, аннотирование, аннулирование, апробирование, аргументирование, ассимиляция, аттестация, балансирование, балансировка, баллотирование, баллотировка, блокирование, бракование, введение, ввод, ввоз, взвешивание, видоизменение, включение, владение, внедрение, внесение, возобновление, воспрещение, воспроизводство, воссоединение, восстановление, востребование, вступление, выборка, выведение, вывоз, выдвигание, выдвигание, выделение, вынесение, выполнение, вытеснение, вычисление, выявление, выяснение, германизация, градирование, группирование, группировка, датирование, дезориентация, декларирование, деление,

демонстрирование, детализация, детализирование, деформирование, дискредитирование, дифференцирование, добавление, добывание, доведение, дозирование, доказывание, документирование, доминирование, дополнение, допуск, дорабатывание, доработка, достижение, дублирование, европеизация, завершение, заготовка, заготовление, загрязнение, задержание, задерживание, задержка, заигрывание, заимствование, заказывание, закладка, закладывание, заключение, закрывание, закрытие, закупка, замалчивание, замедление, замещение, занесение, запаздывание, заполнение, запрет, запрещение, зарождение, засвидетельствование, заслушивание, затрата, зачеркивание, зачисление, защита, злоупотребление, игнорирование, идеализация, идеализирование, идентификация, избежание, избирание, извещение, извлечение, извращение, изготовление, издание, изложение, изменение, измерение, измышление, изображение, изобретение, изолирование, израсходование, изучение, изыскание, иллюстрирование, имитация, имитирование, импровизация, инвентаризация, индивидуализация, инспектирование, инструктирование, интенсификация, информирование, искажение, исключение, искоренение, исполнение, использование, исправление, испытание, исследование, истолкование, истребление, исчезновение, исчисление, калькуляция, канонизация, каталогизация, классификация, классифицирование, кодирование, кодировка, компилирование, комплектация, комплектование, компонование, компоновка, конкретизация, конкурирование, конспектирование, констатация, констатирование, консультирование, контролирование, концентрирование, кооперирование, координирование, копирование, корректирование, культивирование, легализация, лимитирование, локализация, маркирование, маркировка, механизация, моделирование, модернизация, модернизирование, модифицирование, мониторинг, монополизация, монтирование, монтировка, морализация, мотивирование, мотивация, набавка, наблюдение, надбавка, назначение, накапливание, накопление, налаживание, написание, наполнение, направление, настройка, насыщение, начинание, невыполнение, недооценка, недоработка, нейтрализация, нормализация, нормирование, нумерация, нумерование, обвинение, обесмысливание, обеспечение, обеспечивание, обесценение, обесценивание, обзор, обладание, облегчение, обмен, обнаружение, обновление, обобщение, обогащение, обозначение, обоснование, обособление, обострение, обрабатывание, обработка, образование, обращение, обследование, обслуживание, обсуждение, обучение, общение, объединение, объективизация, объявление, объяснение, овладение, оглашение, ограничение, ознакомление, ознаменование, онемечивание, оперирование, описание, определение, опробование, опровержение, опрос, опубликование, организация, ориентирование, освоение, ослабление, осложнение, осмотр, осмысление, осмысливание, оснащение, основание, осознание, оспаривание, остановка, осуждение, осуществление, отбрасывание, отведение, отграничение, отграничивание, отказ, отклонение, открытие, отличие, отметка, отображение, отправка, отправление, отработка, отрицание, отсрочка, отставание, отсчет, отчисление, оформление, очистка, падение, парализация, перевод, перевозка, перевоплощение, перевоспитание,

перевыполнение, перегрузка, перегруппировка, передача, передвижение, переделка, переделывание, переезд, переименование, переквалификация, переключение, перемена, перемещение, перенастройка, перенасыщение, перенесение, перенос, переоборудование, переобучение, переорганизация, переориентация, переориентировка, переосмысление, переоформление, переоценка, перепланировка, перепланирование, переподготовка, переполнение, переработка, перераспределение, перерасход, перерасходование, пересматривание, пересоздание, пересортировка, пересоставление, пересчитывание, переучет, переформирование, переход, перечисление, планирование, повреждение, повторение, повышение, подготовка, подготавливание, поддержание, подерживание, поддержка, подключение, подкрепление, подмена, поднятие, подработка, подразделение, подсчитывание, подтверждение, подчеркивание, подчинение, познание, политизация, получение, понижение, поощрение, пополнение, популяризация, посещение, постановка, постигание, построение, постройка, поступление, потребление, появление, пояснение, правление, превращение, превышение, предложение, предоставление, предохранение, предписание, представление, предупреждение, предъявление, прекращение, премирование, преобразование, преодоление, преподавание, пресечение, преследование, претворение, преувеличение, преуменьшение, преуспевание, прибавление, приближение, приглашение, приготовление, прием, признание, прикрепление, приложение, применение, приобретение, приостановка, приостановление, присвоение, присовокупление, присоединение, приспособление, приуменьшение, приумножение, причинение, причисление, проведение, проверка, провозглашение, прогнозирование, программирование, продажа, продвижение, продление, продолжение, проектирование, произведение, производство, пропагандирование, пропуск, прорабатывание, проработка, просвещение, прослушивание, просматривание, просмотр, противоборство, противопоставление, проявление, прояснение, публикация, пуск, работа, разбивание, разбивка, разбор, разборка, разведение, развитие, раздел, деление, различение, разложение, размен, размещение, размножение, разоблачение, разобщение, разработка, разрешение, разрушение, разъединение, разъяснение, разъединение, разъяснение, распад, распланировка, распознавание, расположение, распоряжение, распределение, распродажа, распространение, распускание, рассказ, расслоение, рассматривание, рассмотрение, рассортировка, рассредоточение, расформирование, расчет, расчленение, расширение, регистрация, регламентация, регламентирувание, регулирование, регулировка, редактирование, резервирование, рекламирование, рекомендация, реконструкция, реформирование, рецензирование, решение, руководство, русификация, саботирование, самоустранение, сбережение, сближение, сбор, сборка, сведение, свержение, сверка, сдача, сдвиг, сепарирование, сжатие, сигнализация, симулирование, синтезирование, систематизирование, скупка, следование, слоение, сложение, слушание, смазывание, смена, смягчение, снабжение, снаряжение, снижение, снятие, снос, сношение, собирание, соблюдение, совершение, совершенствование, совещание, совмещение, согласование, содержание, соединение, создание, созидание,

созревание, сокращение, солидаризация, сообщение, сооружение, соперничество, сопоставление, сопровождение, соревнование, сортирование, сортировка, сосредоточение, составление, соучастие, сохранение, социализация, сочетание, сочинение, сочленение, спад, спасание, спасение, специализация, списание, сплачивание, сплетение, спуск, сравнение, сравнивание, сращение, сращивание, срезание, ссылка, стимулирование, стирание, столкновение, страхование, стыкование, стягивание, стяжка, субсидирование, сужение, суммирование, сформирование, схематизация, сцепление, счет, съезд, таксировка, теоретизирование, тестирование, течение, типизирование, толкование, торможение, транспортировка, трансформирование, убавление, убеждение, убывание, убыстрение, увековечение, увеличение, увлажнение, увольнение, угасание, углубление, угнетение, удаление, удержание, удешевление, удлинение, удовлетворение, удостоверение, ужесточение, узаконение, указание, уклонение, укомплектование, укорачивание, укоренение, укорочение, укрепление, укрупнение, укрытие, улаживание, уличение, улучшение, уменьшение, умножение, умолчание, унификация, уничтожение, уплата, упоминание, употребление, управление, упражнение, упразднение, упрощение, уравнение, уравнивание, урегулирование, усвоение, ускорение, усложнение, усовершенствование, установка, установление, устранение, устройство, усугубление, уточнение, утрирование, утяжеление, ухудшение, уценка, учеба, учение, учет, учреждение, уяснение, фабрикация, фальсифицирование, фигурирование, фиксация, фиксирование, фильтрация, фильтрование, фильтровка, финансирование, формирование, формулирование, форсирование, функционирование, ходатайство, хозяйствование, хранение, хронологизация, хронометрирование, централизация, циркуляция, цитирование, чередование, членение, чтение, шифрование, штампование, эволюционирование, эксплуатирование, экспонирование, экспортирование, элиминирование, юстирование};

Властивість

$N(m_6, ls_2) =$ {абсолютность, автономность, авторитарность, авторитетность, агрессивность, адекватность, азартность, академичность, аккуратность, активность, актуальность, аллегоричность, алогичность, альтруистичность, аляповатость, аморальность, аналогичность, анахроничность, аномальность, апатичность, аполитичность, апостериорность, априорность, аристократичность, артистичность, архаичность, асимметричность, атрибутивность, аутентичность, афористичность, банальность, баснословность, безалаберность, безапелляционность, безграничность, бездарность, бездоказательность, безличность, безмерность, безнадежность, безнаказанность, безнравственность, безобидность, безоговорочность, безответственность, безотчетность, безошибочность, безрезультатность, безукоризненность, безусловность, безуспешность, безынициативность, бесконечность, бесконтрольность, бесконфликтность, бескорыстность, бесперебойность, бесперспективность, бесплановость, бесплодность, бесполезность, беспорядочность, беспочвенность, беспредельность, беспредметность, беспрекословность, беспрепятственность,

беспрерывность, беспримерность, беспринципность, беспристрастность, беспричинность, бессистемность, бесславность, бессменность, бессмертность, бессмысленность, бессознательность, бесспорность, бессрочность, бесформенность, бесцельность, бесцензурность, бесчисленность, благотворность, буквальность, важность, верность, весомость, вещественность, виртуозность, внеочередность, внеплановость, вразумительность, всеобщность, выполнимость, гармоничность, гениальность, гибкость, гиперболичность, гипотетичность, годность, голословность, готовность, грамотность, грандиозность, громадность, гуманность, двойственность, двусмысленность, двусторонность, действительность, декларативность, дельность, детальность, дефективность, дефицитность, динамичность, дипломатичность, длительность, добротность, догматичность, доказуемость, доскональность, дословность, достаточность, достоверность, доступность, доходчивость, драматичность, единичность, емкость, естественность, желательность, живописность, жизнеспособность, жизнестойкость, загадочность, заинтересованность, законность, закономерность, законченность, закоренелость, замедленность, заметность, замкнутость, замысловатость, занимательность, заостренность, запоздалость, запоминаемость, запутанность, запущенность, затейливость, затрудненность, затруднительность, заурядность, зачаточность, знаменательность, знаменитость, идеальность, идейность, идентичность, избирательность, избыточность, известность, извечность, извращенность, изменчивость, измеримость, изначальность, изолированность, изоляция, изошренность, изученность, изысканность, иллюзорность, иллюстративность, имманентность, импульсивность, инвариантность, индифферентность, индуктивность, интеллектуальность, интенсивность, информативность, иррациональность, исконность, искусственность, исполнимость, исправность, истинность, историчность, каноничность, капиталность, категоричность, качественность, книжность, коллегияльность, колоритность, колоссальность, компетентность, компилятивность, комплексность, комплектный, конечность, конкретность, консервативность, конспективность, конструктивность, контрастность, концентрированность, корпоративность, корректность, красноречивость, красочность, кратковременность, краткосрочность, краткость, критичность, кропотливость, курьезность, лаконичность, легальность, легендарность, лживость, либеральность, литературность, логичность, ложность, локальность, лояльность, льготность, максимальность, маловероятность, малодоступность, малозначительность, малоупотребительность, малочисленность, массивность, массовость, масштабность, материальность, машинальность, мгновенность, методичность, механичность, минимальность, мнимость, многогранность, многозначительность, многозначность, многократность, многоплановость, многосложность, многостепенность, многосторонность, многоступенчатость, многочисленность, множественность, монолитность, монопольность, монументальность, надежность, назидательность, налаженность, направленность, напряженность, настоятельность, насущность, натуральность, наукообразность, неаккуратность, неактуальность, неблагозвучность, небрежность, невероятность, невещественность, невнимательность,

невозвратимость, невосполнимость, невразумительность, невыгодность,
невыполнимость, невыразительность, негодность, неграмотность,
недальновидность, недвусмысленность, недействительность, неделимость,
недоброкачественность, недоказуемость, недолговечность, недопустимость,
недостаточность, недостижимость, недостоверность, недоступность,
недосягаемость, неестественность, нежелательность, независимость,
незаинтересованность, незаконность, незаконченность, незаменимость,
незамысловатость, незанимательность, незатейливость, незначительность,
неизбежность, неизвестность, неизменность, неизменяемость, неизмеримость,
неизученность, неисполнимость, неисправимость, неисчерпаемость,
нейтральность, некомпетентность, некомплектность, некорректность,
некритичность, нелегальность, нелогичность, нелояльность, немаловажность,
немногословность, немногочисленность, ненадежность, ненормальность,
ненужность, необдуманность, необозримость, необоснованность,
необработанность, необратимость, необъективность, необъяснимость,
необъятность, необычайность, необычность, необязательность, неограниченность,
неоднородность, неожиданность, неоконченность, неопределенность,
неопределимость, неопровержимость, неорганизованность, неосведомленность,
неосмотрительность, неосновательность, неосознанность, неоспоримость,
неотвратимость, неотделимость, неотложность, неотчетливость,
непозволительность, неполноценность, непонятность, непоправимость,
непоследовательность, непосредственность, неправильность, неправомерность,
неправомочность, непрактичность, непреднамеренность, непредотвратимость,
непредусмотрительность, непреодолимость, непререкаемость, непрерывность,
непривычность, непригодность, неприемлемость, неприкосновенность,
неприменимость, неприспособленность, непричастность, непродолжительность,
непродуктивность, непроизводительность, непроизвольность,
непропорциональность, непрочность, неравномерность, неравноправность,
неразвитость, неразгаданность, нераздельность, неразложимость, неразрешимость,
неразрывность, неразумность, нерасчленимость, нерациональность, нереальность,
нерегулярность, нерентабельность, неритмичность, нерушимость,
несамостоятельность, несбыточность, несвоевременность, несвязность,
нескладность, несменяемость, несовместимость, несогласованность,
несоизмеримость, несомненность, несообразность, несоразмерность,
несостоятельность, несправедливость, нестандартность, нестройность, неточность,
неубедительность, неукоснительность, неумеренность, неупотребительность,
неустойчивость, неустроенность, неутешительность, неуязвимость,
нецелесообразность, нечеткость, неясность, ничтожность, новизна, нормативность,
нравоучительность, нужность, обдуманность, обильность, обманчивость,
обобщенность, обоснованность, обособленность, обостренность, обоюдность,
образность, обратимость, обрывистость, обстоятельный, общедоступность,
общеизвестность, общераспространенность, общеупотребительность,
объективность, обязательность, ограниченность, односторонность, одновременность,
однозначность, однократность, однообразность, однородность, односторонность,

оперативность, определенность, определимость, оптимистичность, опытность, организованность, органичность, ординарность, оригинальность, ортодоксальность, осмысленность, основательность, осознанность, отвлеченность, относительность, отчетливость, официальность, официозность, очевидность, ошибочность, ошутимость, парадоксальность, парность, педагогичность, педантичность, первичность, первозданность, первоочередность, переменчивость, перенасыщенность, переходность, периодичность, перспективность, пессимистичность, пикантность, планомерность, плачевность, поверхностность, повседневность, повсеместность, повторность, подверженность, подвижность, поддельность, подлинность, подложность, подробность, подсознательность, подчеркнутость, пожизненность, позитивность, познаваемость, показательность, полемичность, полновесность, полноправность, положительность, понятность, популярность, поразительность, портативность, последовательность, посредственность, постепенность, постоянство, потенциальность, поточность, похожесть, почетность, правдивость, правдоподобность, правильность, правомерность, правомочность, практичность, предварительность, предвзятость, предельность, предметность, преднамеренность, предопределенность, предположительность, предпочтительность, прееменность, преждевременность, преступность, претенциозность, приближительность, прибыльность, приверженность, привлекательность, пригодность, приемлемость, призрачность, примерность, примечательность, примитивность, принципиальность, приоритетность, притягательность, проблематичность, проблемность, прогрессивность, продолжительность, продуктивность, продуманность, прозрачность, произвольность, пропорциональность, просторность, простота, пространность, противоестественность, противозаконность, противоположность, противоречивость, профессиональность, прямолинейность, прямота, публичность, равнозначность, равномерность, равноценность, радикальность, развитость, разительность, разлаженность, разнообразность, разносторонность, разность, разнотипность, разобщенность, разорительность, разрозненность, разумность, разъединенность, разумность, разъединенность, расплывчатость, распространенность, рассудочность, рассчитанность, растянутость, рациональность, реакционность, реалистичность, реальность, революционность, регрессивность, регулярность, резонность, результативность, решительность, рискованность, риторичность, родственность, сакраментальность, самобытность, самокритичность, самопроизвольность, самостоятельность, сатиричность, свертываемость, сверхъестественность, светопроницаемость, светостойкость, светочувствительность, своевременность, своеобразность, связанность, связность, сдержанность, секретность, сенсуальность, сентениозность, сепаратность, серьезность, сжатость, символичность, симметричность, синхроничность, систематичность, скандальность, скачкообразность, скептичность, складность, скоропостижность, скороспелость, скрупулезность, скудность, скученность, слаженность, слитность, сложность, смежность, смертоносность, смешанность, смутность, собирательность, совместность, современность, согласованность,

содержательность, сознательность, соизмеримость, сокрушительность, солидность, сомнительность, сообразность, соотносительность, сопредельность, сопротивляемость, сохранность, сочетаемость, спасительность, специальность, специфичность, сплоченность, спорность, справедливость, спутанность, срочность, стабильность, стандартность, стародавность, старозаветность, старомодность, старорежимность, статичность, стереотипность, стилизованность, стильность, стихийность, стойкость, странность, стремительность, строгость, стройность, ступенчатость, субстанциональность, субъективность, суверенность, сумбурность, существенность, сходность, схоластичность, таинственность, текстуальность, текучесть, телесность, теоретичность, техничность, технологичность, типичность, тлетворность, тождественность, торжественность, тотальность, точность, трагичность, традиционность, транспортабельность, трансцендентальность, трансцендентность, трафаретность, тренированность, тривиальность, трудность, трудоемкость, тщательность, тщетность, тяжесть, убедительность, увлекательность, углубленность, удобность, удобоисполнимость, удобопонятность, удобочитаемость, удобство, удовлетворительность, узость, уклончивость, ультимативность, умеренность, уместность, умышленность, универсальность, упорядоченность, употребительность, упрощенность, усвояемость, условность, успешность, устойчивость, утешительность, утилитарность, утонченность, уязвимость, фальшивость, фанатичность, фантастичность, фаталистичность, фатальность, феноменальность, фигуральность, фиктивность, формалистичность, формальность, фрагментарность, фундаментальность, функциональность, хаотичность, характеристичность, характерность, химеричность, художественность, целесообразность, целеустремленность, целостность, цельность, цензурность, цивилизованность, цикличность, частичность, частность, частота, человечность, четкость, членораздельность, чрезвычайность, чрезмерность, чудодейственность, чуждость, шаблонность, широта, щепетильность, эквивалентность, экзотичность, эклектичность, экономичность, экономность, экспансивность, экспрессивность, экстенсивность, экстравагантность, экстраординарность, экстренность, эластичность, элементарность, эмпиричность, энциклопедичность, эпизодичность, эфемерность, эффektivность, эффektivность, явственность, яркость, ясность }.

ЗМІСТ

ВСТУП	3
I. ОГЛЯД СУЧАСНИХ ДОСЛІДЖЕНЬ У ГАЛУЗІ АВТОМАТИЧНОГО РЕФЕРУВАННЯ	9
1.1. Розвиток ідеї автоматичного реферування	9
1.2. Методи автоматичного реферування	20
1.3. Огляд лінгвістичних систем автоматичного реферування	30
1.4. Моделювання знань у системах автоматичного реферування	39
Висновки	53
II. ПОБУДОВА СИСТЕМИ АВТОМАТИЧНОГО РЕФЕРУВАННЯ З ОПЕРТЯМ НА ЗНАННЯ.....	54
2.1. Моделювання компресії тексту в процесі реферування.....	54
2.1.1. Компресія на морфологічному рівні	60
2.1.2. Компресія на синтаксичному рівні	63
2.2. Моделювання компресії змісту тексту	71
2.3. Експериментальна перевірка роботи алгоритму автоматичного реферування.....	84
Висновки	95
III. МОДЕЛЮВАННЯ СЕМАНТИЧНИХ ЗВ'ЯЗКІВ «ТЕКСТ-РЕФЕРАТ» ІЗ ВИКОРИСТАННЯМ ОНТОЛОГІЙ	97
3.1. Побудова семантико-контекстної моделі реферування.....	97
3.2. Обґрунтування вибору онтологій як засобу подання знань у системі автоматичного реферування та мови їх опису.....	106
3.3. Створення онтологій для системи автоматичного реферування	117
Висновки	126
ПІДСУМОК	129

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	132
<i>Додаток А</i>	159
<i>Додаток Б</i>	164
<i>Додаток В</i>	166
<i>Додаток Г</i>	167
ЗМІСТ.....	178

Наукове видання

Лазаренко Ольга Володимирівна

Панченко Дмитро Ігорович

**ДОСЛІДЖЕННЯ СЕМАНТИЧНИХ ЗВ'ЯЗКІВ
ТЕКСТ-РЕФЕРАТ
ДЛЯ СТВОРЕННЯ СИСТЕМИ АВТОМАТИЧНОГО РЕФЕРУВАННЯ З
ОПЕРТЯМ НА ЗНАННЯ**

Монографія

Відповідальна за випуск *О. В. Лазаренко*

Редактор *З. М. Москаленко*

Комп'ютерний набір *Д. І. Панченко*

Підписано до друку 21.01.2013. Формат 60×84/16.

Папір офсетний. Гарнітура «Times».

Ум. друк. арк. 10,46. Обл.-вид. арк. 9,87.

Тираж 100 пр. Зам. № 22/12.

План 2012/13 навч. р., поз. №5 у переліку робіт кафедри

Видавництво

Народної української академії

Свідоцтво № 1153 від 16.12.02.

Надруковано у видавництві

Народної української академії

Україна, 61000, м. Харків, МСП, вул. Лермонтовська, 27.